The Dissertation Committee for Megan Kathleen Mulligan Certifies that this is the approved version of the following dissertation:

TRANSCRIPTOMICS AND THE GENETICS OF ALCOHOL CONSUMPTION IN MICE

Committee:

Susan E. Bergeson, Co-Supervisor

R. Adron Harris, Co-Supervisor

Viswanath R. Iyer

Orly Alter

Seema Agarwala

Edward M. Marcotte

TRANSCRIPTOMICS AND THE GENETICS OF ALCOHOL CONSUMPTION IN MICE

by

Megan Kathleen Mulligan, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Dedication

To my parents, Steve and Susan Mulligan, for their constant love and support.

To my family and friends, whose words of encouragement helped to elevate me above despair during the long and winding road of graduate school.

And, especially, to Grandma Alice and Grandpa Bud and Grandma Iris and Grandpa Herb, who will be my role models forever.

Acknowledgements

I would first like to thank my supervisors, Dr. Susan E. Bergeson and Dr. R. Adron Harris, for having faith in me and for allowing me the opportunity to study science under their tutelage. I would especially like to thank Dr. Bergeson for allowing me to work in her laboratory and for giving me every possible opportunity to grow as a scientist and I would like to additionally thank all the former members of the Bergeson Laboratory for their help and support along the way, especially; Ari Berman, Pat Levin, Kerrie Lodowski, Julie Owen and Gwen Schafer. Microarray analysis would not have been possible at the University of Texas without the help and support of Dr. Vishwanath Iyer, Patrick Killian, David Parr and Elizabeth Osterndorff-Kahanek. I would like to thank M. June and J. Virgil Waggoner whose generous contribution in the wake of great tragedy made possible the Waggoner Center for Alcohol and Addiction Research. I would also like to thank all of the members of the Waggoner Center, especially; Debra C. James, Dr. R. Dayne Mayfield and Dr. Igor Ponomarev. I would also like to extend my thanks to Dr. William E. Cassady, for providing me with a place to write this dissertation.

TRANSCRIPTOMICS AND THE GENETICS OF ALCOHOL CONSUMPTION IN MICE

Publication No._____

Megan Kathleen Mulligan, Ph.D.

The University of Texas at Austin, 2007

Supervisors: Susan E. Bergeson and R. Adron Harris

Alcoholism is a complex disease determined by both genetic and environmental components that exerts a devastating economic and social impact worldwide. The complexity of this disease makes the elucidation of candidate genes for the susceptibility to alcoholism difficult in human populations, however, mouse model systems replicate many aspects of the disease and represent an excellent system for the investigation of the genetic contributions to alcoholism. One component of alcoholism that can be investigated in mouse models is the predisposition to high alcohol consumption. Selectively bred and inbred mice differ markedly in the level of voluntary alcohol intake using a two-bottle choice paradigm. The phenotype of voluntary alcohol consumption in mice is a complex trait and a genetic comparison between mouse models with similar levels of alcohol intake should identify genes that contribute to the predisposition for alcohol consumption. Three different studies were completed at the University of Texas and candidate genes involved in the predisposition to high alcohol consumption in mice

were identified through the use of brain transcriptome analysis. In the first study, 3,800 transcripts were identified that were divergent between 3 selected lines and 6 isogenic strains of mice known to differ in voluntary alcohol consumption. This list was filtered to reveal candidate genes associated with alcohol preference on mouse chromosome 9: *Arhgef12, Carm1, Cryab, Cox5a, Dlat, Fxyd6, Limd1, Nicn1, Nmnat3, Pknox2, Rbp1, Sc5d, Scn4b, Tcf12, Vps11, Zfp291*. In the second study, analysis of voluntary alcohol intake and brain gene expression between two closely related inbred mouse substrains separated for nearly fifty years revealed divergent alcohol consumption as well as genetic variation between the substrains. Finally, the third study revealed dominant and overdominant patterns of expression in an F1 hybrid that voluntarily consumed more alcohol than either inbred parental strain. The microarray datasets analyzed here represent an important first step in the elucidation of the genetic determinants of high alcohol consumption in mice and will be influential in the discovery of genes that play a role in vulnerability to alcoholism in humans.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

ALCOHOL, ADDICTION AND HUMANITY

Humankind has had a long love affair with ethyl alcohol. Ethanol, in the form of beer and wine and much later in the form of distilled spirits, has been consumed by human civilizations since at least 4,000 B.C. and archaeological evidence suggests that Neolithic humans, circa 10,000 B.C., also consumed fermented beverages (Hansen 1995). Until the introduction of distillation methods in about 700 A.D., most alcoholic beverages were of low ethanol content and were probably consumed as a safe alternative fluid to water, which was likely contaminated by disease and dangerous to consume (Vallee 1998). By the 19th century, highly concentrated forms of ethanol were cheap and widely available and alcoholism was being recognized as a serious and pervasive disease.

Today, alcoholism affects millions of people worldwide and continues to be a serious problem in America. Alcohol abuse and dependence in this country contributes to nearly 100,000 deaths annually (Mokdad, Marks et al. 2004) and drains over $150 billion a year in alcohol related costs (Harwood, Fountain et al. 1999). Over half of all Americans know of a family member who has suffered or is suffering from alcohol dependence (Dawson and Grant 1998). Worldwide, as many as 1 in 650 children are born with FAS (Fetal Alcohol Syndrome), a condition arising from exposure to alcohol in the womb that can lead to cognitive deficits, behavioral problems and facial deformities (Olegard, Sabel et al. 1979).

Currently, there is no common cure for alcohol addiction. Although many treatment options exist, it is not yet possible to predict if a given treatment will be successful and the results from different therapies vary from patient to patient. Despite

variability in success rates, the most common form of treatment for alcohol dependence in the United States is still group counseling (McLellan, Carise et al. 2003). A few drugs that take advantage of the pharmacology of alcohol on metabolic enzymes, disulfiram, and endogenous brain neurotransmitter systems, acamprosate and naltrexone, have been approved for the treatment of alcoholism in the United States and other similarly designed drugs have shown promise in clinical trials. However, to date, no single drug therapy for alcoholism has proven efficacious in successfully treating all populations of alcoholics. The tenuous link between disease state and treatment efficacy is likely due to the wide range of targets affected by alcohol coupled with the complex nature of the disease.

MOLECULAR TARGETS OF ALCOHOL

Alcohol is a small and promiscuous molecule that exerts an effect in nearly all the organ systems of the body and has a wide range of targets within the nervous system. Initially, alcohol acts as a nervous system depressant that enhances inhibitory GABAergic transmission through potentiation of $GABA_A$ receptors and reduces excitatory glutamatergic transmission by inhibition of NMDA receptors (Eckardt, File et al. 1998) but alcohol also has an effect on many other neurotransmitter systems in the brain including the dopamine, serotonin, opioid and cholinergic pathways (Mereu, Fadda et al. 1984; Gessa, Muntoni et al. 1985; Campbell, Kohl et al. 1996; Cardoso, Brozowski et al. 1999; Oswald and Wand 2004). Alcohol is also known to influence the activity of several 2^{nd} messenger systems and transcriptional regulators including up-regulation of cAMP through CREB activation and activation of ΔFosB (Nestler 2005). Other targets of alcohol include the stress pathway, which is targeted through activation of the hypothalamic-pituitary-adrenal (HPA) axis and through regulation of neuropeptides such as corticotrophin-releasing factor and neuropeptide Y (Cowen, Chen et al. 2004). Alcohol has an effect on many other brain proteins far too numerous to describe in detail here, but it should be evident that alcohol has a profound effect on many systems in the brain including the reward centers of the brain.

REWARD PATHWAYS AND THE DEVELOPMENT OF ADDICTION

Drugs of abuse (heroine, cocaine, ethanol, nicotine, etc.) have various molecular targets in the brain and differ in their chemical composition but they all share a common defining feature; drugs of abuse target the brain's reward system (Wise 1998; Nestler 2005). In the 1950's it was discovered that electrical stimulation of certain brain regions had a rewarding effect in animals and this stimulation acted as stronger positive reinforcement for behavior than natural rewards like food (Olds and Milner 1954). These initial observations led to the mapping of the brain regions and neural circuits involved in the reward system.

The mesolimbic dopamine circuit is the central processor of the reward system. It includes dopamine neurons whose cell bodies reside in the ventral tegmental area (VTA) of the midbrain and send projections to the nucleus accumbens (NAc) and other targets in the limbic forebrain where they release the neurotransmitter dopamine (figure 1.1). This basic reward circuit modulates and is modulated in return by other brain regions and other neurotransmitter systems. Multiple neurotransmitter systems play a role in the mesolimbic dopamine reward circuit including GABAergic, glutamatergic, opioid, serotinergic and cholinergic neurotransmitter systems. Many different brain regions implicated in addiction are connected to the reward pathway as well, including the frontal cortex, hippocampus, amygdala and the hypothalamus. The reward pathway is an evolutionarily ancient brain system that is thought to mediate survival by reinforcing critical behaviors such as food intake and sex. Nearly all drugs abused by humans that are self-administered by animals activate the reward circuitry of the brain and lead to initial increases in dopamine levels in the NAc (Wise and Rompre 1989; Dackis and O'Brien 2005). Alcohol, because it has an effect on most of the

neurotransmitter systems in the reward pathway and many diverse targets in the brain, can influence the reward pathway through multiple mechanisms (figure 1.2)

Whether the reward center is being activated by natural rewards or drugs of abuse, stimulation of the reward pathway causes a release of dopamine in the reward centers of the brain which induces a positive sensation of euphoria. Through positive reinforcement, drug intake becomes associated with pleasant sensations contributing to further drug use. Both genetic and environmental factors can interact to produce individual variation in the initial response to a drug and each drug of abuse has a unique effect on the reward system. In individuals with genetic or environmental risk factors, initial exposure to a drug may seem more pleasurable or have less aversive effects. These risk factors could contribute to repeated drug administration and to loss of control over drug intake, the hallmark of drug addiction. Over time, repeated exposure to a drug leads to changes in gene expression as well as changes in the neuronal architecture of the brain, referred to as synaptic plasticity or neuroadaptation. Neuroadaptation is part of a normal homeostatic mechanism that occurs as the brain adjusts to the presence of the drug and makes compensatory changes. It has been hypothesized that repeated disruption of the reward centers or stress pathways of the brain can lead to pathway dysregulation. The allostatic model of addiction proposes that, in response to continuous drug induced dysregulation, the brain adjusts to a new stable set point that is functional but pathological (Koob 2003). The new, drug adjusted set point could contribute to negative dysphoric states such as tolerance, withdrawal and craving that promote drug abuse and addiction. Figure 1.3 details the mechanism of progression to addiction. Many theories of addiction exist that attempt to describe the major transition from casual drug use to compulsive drug use, for a review on neuropsychological theories of drug addiction, see (Le Moal and Koob 2007). A major goal for the treatment and prevention of alcoholism

has been to define the genetic components that contribute to the genetic risk for or the predisposition to addiction to alcohol.

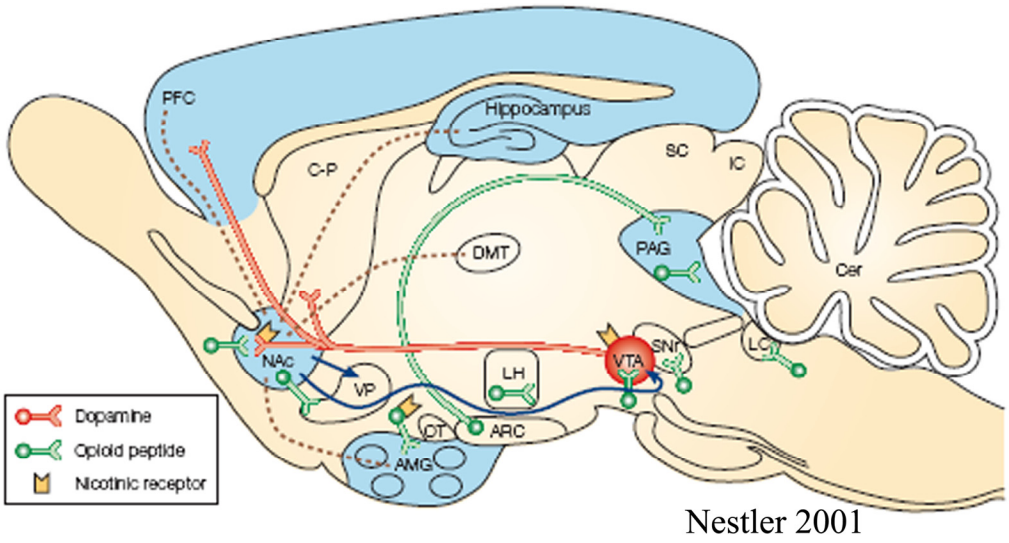


Figure 1.1: Reward circuits stimulated by drugs of abuse. The dotted lines indicate afferent pathways to the nucleus accumbens (NAc). The blue lines indicate efferent pathways from the nucleus accumbens. The blue shading indicates the location of $GABA_A$ receptor complexes. ARC, arcuate nucleus; Cer, cerebellum; DMT, dorsomedial thalamus; IC, inferior colliculus; LC, locus coeruleus; LH, lateral hypothalamus; PAG, periaqueductal grey; SC, superior colliculus; SNr, sbstantia nigra pars reticulatal; VP, ventral pallidum (Nestler 2001).

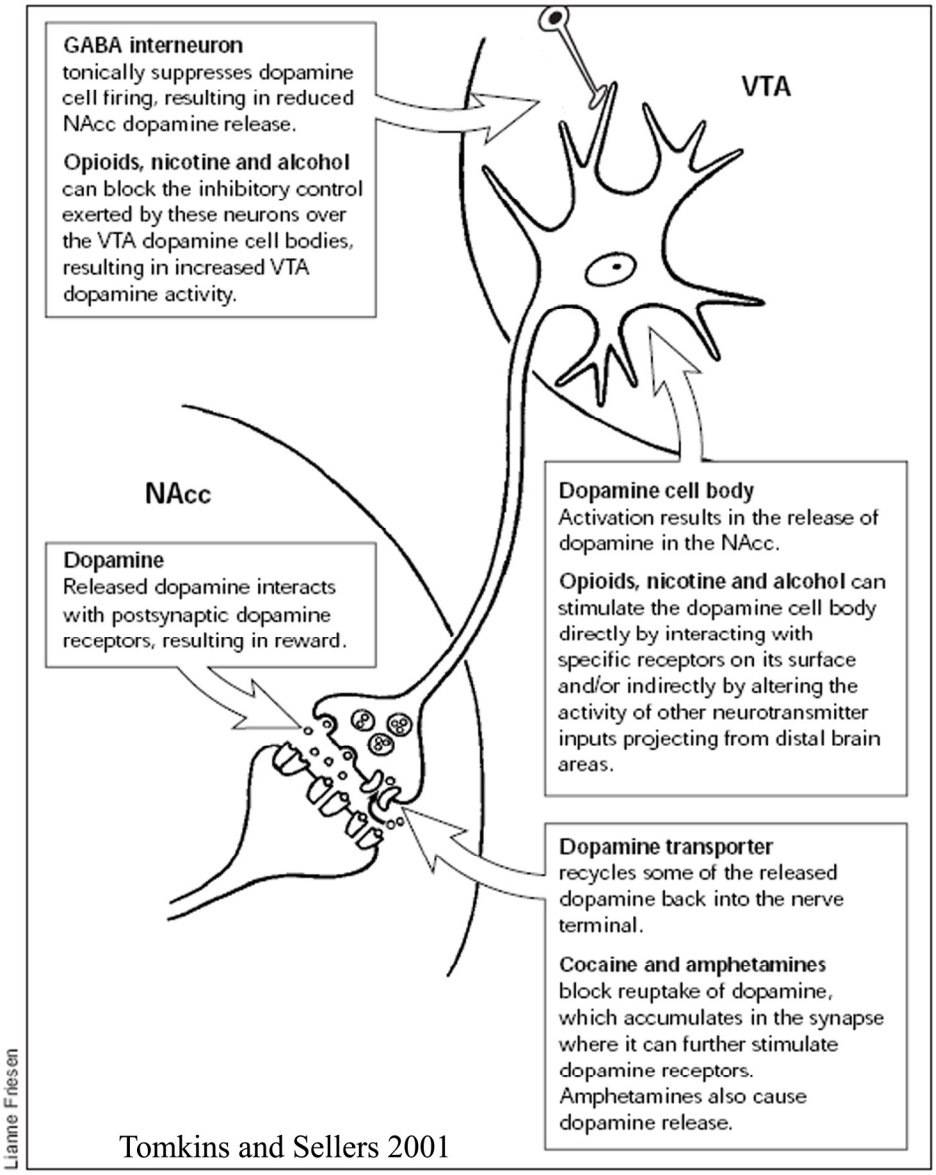
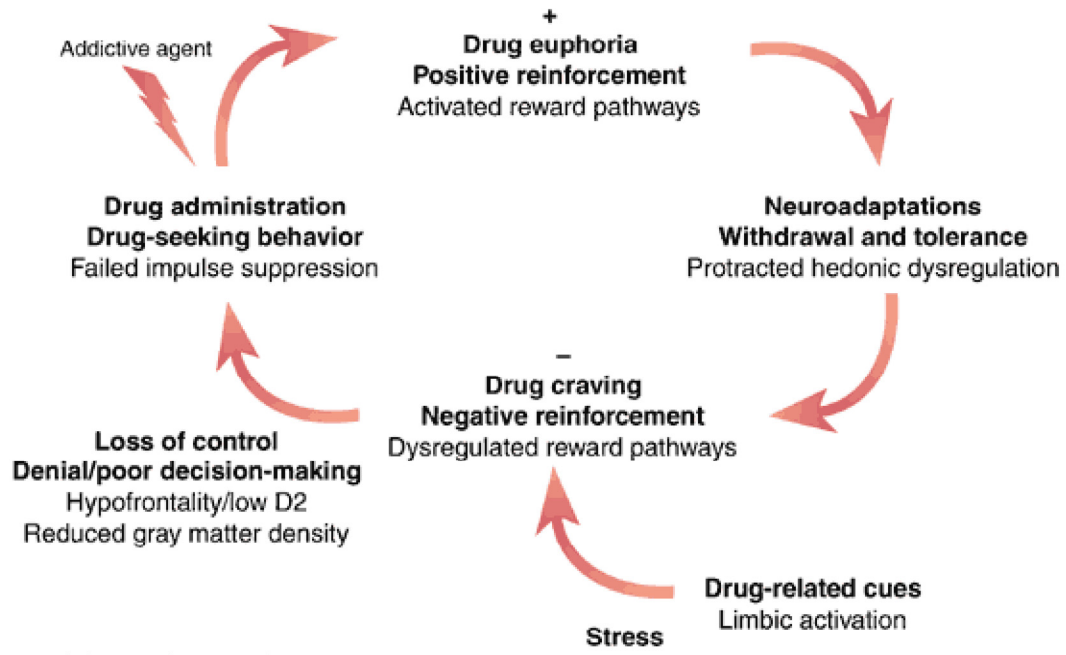


Figure 1.2: The effects of alcohol and other drugs on dopamine release. Dopaminergic neurons in the ventral tegmental area (VTA) send projections to the nucleus accumbens (NAcc) where dopamine is released. Alcohol, nicotine and opioids have multiple targets within this circuit where they can act to modulate dopamine release (Tomkins and Sellers 2001).



Dackis and O'Brien 2005

Figure 1.3: Addiction cycle. Administration of a drug leads to activation of the reward center. Continued drug use can lead to changes in the brain that compliment continued drug use. Addiction is characterized by a lack of control over drug use. Genetic and environmental factors, although not shown in the figure, contribute to each step in the cycle (Dackis and O'Brien 2005).

THE HUMAN GENETICS OF ALCOHOLISM

Alcoholism is a complex disease determined by many genetic and environmental factors. Twin studies in humans have estimated that 30-60% of the risk of alcohol addiction is inherited (Walters 2002; Goldman, Oroszi et al. 2005; Hiroi and Agatsuma 2005). Detecting the genes that contribute to addiction is a daunting task; alcoholism is a heterogenous disease with multiple phenotypes likely influenced by many genes of small to moderate effect size that interact with environmental factors to produce the final disease state. As a result of the complexity of the disease, it has been extremely difficult to associate genetic differences between human populations of alcoholics and non-alcoholics with vulnerability for addiction.

The primary method for detecting genes that contribute to alcohol related phenotypes in human populations has been linkage and association studies. Both methods compare the frequencies of genetic polymorphisms between different human populations such as alcoholics and non-alcoholics. However, linkage analysis consists of whole genome scans for regions of interest that contain genes contributing to addiction with no *a priori* assumptions about the location of the genetic polymorphisms whereas association studies limit the analysis to specific candidate genes hypothesized to be involved in the phenotype of interest. Although these types of analyses have proven useful for the identification of candidate genes involved in monogenic diseases for which a single loci exerts a large effect, the large number of small effect size genes involved in alcoholism has made gene identification difficult (Enoch 2003). In order to detect these small effect genes, large sample sizes and well defined populations are required to perform these types of analyses. The presence of confounding factors, such as the comorbidity of alcoholism with other psychiatric disorders and general substance abuse, can lead to conflicting results. Despite the existence of large and carefully phenotyped

datasets, including families of alcoholics and non-alcoholics from the Collaborative Study on the Genetics of Alcoholism (Begleiter H 1995), relatively little progress has been made in definitively identifying genes associated with the risk for alcoholism. A recent review of alcoholism association studies revealed that, after decades of research, only a handful of genes are reliably associated with alcoholism in different human populations of alcoholics; these include genes involved in the metabolism of alcohol, ADH and ALDH, and GABA neurotransmission (Kohnke 2007). Tentative associations have been found for many genes involved in synaptic transmission and other neurobiological processes but the significance or the direction of the association of these genes with the addiction phenotype often varies depending on which populations are being tested (Kohnke 2007). Genome wide linkage studies in humans using dense marker mapping have identified several regions of interest for the genetic involvement in alcohol and addiction, but further testing will be required to identify the specific genes that contribute to alcoholism within each chromosomal region (Johnson, Drgon et al. 2006; Liu, Drgon et al. 2006; Agrawal, Hinrichs et al. 2007).

Alcoholism likely manifests in humans based on the interaction of different environmental circumstances, such as age of onset or childhood trauma, with various subsets of genes that contribute to different phenotypes, such as personality traits, stress responsivity or mental illness (Kreek, Nielsen et al. 2005). The complexity of these interactions probably produces many different populations of alcoholics and each population may share common genetic risk factors whose effect size is minimized beyond detection when different populations are combined and analyzed in linkage or association studies. The use of intermediate phenotypes (endophenotypes) associated with alcoholism that discriminate between subgroups is one attempt that has been made to sort through the tangled mess of genetic and environmental issues that convolute our

understanding of the genetics of alcohol abuse and addiction (Schuckit 2000). A reduction in complexity is a necessary requirement for improving our understanding of this devastating disease and the enlistment of animal models that mirror different aspects of the disease state in humans has greatly increased our understanding of the genetic components of alcoholism.

GENETIC ANIMAL MODELS

Alcoholism is a human condition, but animals as diverse as our fellow primates and the invertebrate fruit fly have been used to model different components of alcoholism. Non-human primates are intriguing animal models with regards to addiction biology because of the genetic, behavioral and neuroanatomical similarities to humans. Ethical issues aside, there are many disadvantages to working with non-human primates that limit their widespread use as a genetic model system for studying alcohol related behaviors including cost, a long reproductive cycle and limited genetic techniques (Barr and Goldman 2006). Invertebrate systems such as *Drosophila melanogaster,* are cheap to maintain, have short life cycles and offer an excellent genetic model system. The fruit fly has been used to investigate the genetic components of alcohol tolerance (Ghezzi, Al-Hasan et al. 2004; Scholz, Franz et al. 2005) and sensitivity to alcohol (Corl, Rodan et al. 2005; Rothenfluh, Threlkeld et al. 2006) but it is difficult to recapitulate some of the more complicated aspects of the human disease such as motivation to drink, craving and relapse. Based on cost, speed, genetic utility and the ability to replicate many different aspects of alcoholism, rodent models have been the research organism of choice in the alcohol field for over half a century.

Rodent models

The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) describes several human behaviors that are used in the diagnosis of alcoholism: tolerance, increased amounts are needed to achieve the desired effect; the presence of withdrawal symptoms after cessation of drinking; consuming more alcohol than intended; desire or unsuccessful attempts to reduce drinking; spending a large amount of time recovering or acquiring alcohol; drinking interferes with social or occupational duties; and continued use of alcohol despite the detrimental effects of

continued drinking (American Psychiatric Association 1994). While these diagnostic criteria do not capture all the intricacies of human related behavior, they have been used to judge the validity of animal models of addiction (Tabakoff and Hoffman 2000). No rodent model is expected to capture all of the elements of the human disease, but rat and mouse genetic model systems do replicate different aspects of the addiction process, such as tolerance and withdrawal. Rodents are also excellent genetic models for human diseases because the rodent and human genomes contain many syntenic regions (Mural, Adams et al. 2002). Rodent model systems also demonstrate voluntary self administration of alcohol which is an extremely important component for studying alcoholism because drug addiction is always preceded by repeated voluntary drug intake.

Voluntary alcohol consumption

Many different experimental procedures have been developed in which rats and mice will voluntarily self administer solutions of alcohol (Tabakoff and Hoffman 2000; Spanagel 2003). The simplest of these techniques is two-bottle choice alcohol consumption. In this paradigm rodents are offered a choice of either water or an alcohol solution. Most rodents find low concentrations of alcohol (roughly less than 6% alcohol and water) appealing and will readily prefer to drink alcohol over water. At higher concentrations (most studies use 10% to 15% concentrations of alcohol to water) there is variability in the level of alcohol each rodent model consumes. Outbred rodent population with heterogeneous genetic backgrounds show individual variation in the amount of alcohol consumed indicating that alcohol consumption is a quantitative trait. This paradigm also measures preference for alcohol; a ratio greater than 50% alcohol volume to total volume consumed indicates a preference for the alcohol solution. Many different studies have characterized the level of alcohol preference for and the voluntary consumption of different concentrations of alcohol using two-bottle choice paradigms in

different rodent models. The level of intake varies in a quantifiable and highly replicable manner between different genetic model systems, across laboratories and even over time (Wahlsten, Bachmanov et al. 2006). Two-bottle choice consumption of alcohol does not directly measure the motivation to drink or the subjective rewarding properties of alcohol in the rodent model system but it is a genetically determined trait and the genes involved in high alcohol consumption in rodents are likely candidate genes for the contribution to addiction vulnerability in humans.

Selectively Bred Lines

Selective breeding takes advantage of population heterozygosity. Two-bottle choice alcohol consumption is a quantitative trait, so in a normally distributed population there will be a range of individuals that vary between high and low alcohol consumption. Over many generations, selective breeding of high and low responders lengthens the magnitude of the drinking response leading to a high alcohol drinking line and a low alcohol drinking line. If selection begins with a heterozygous (outbred) population and extensive inbreeding is avoided, the resulting high and low selectively bred lines should capture genes responsible for the level of consumption and unrelated genes will be randomly distributed in each line.

Several rat and mouse selectively bred lines have been created based on high and low levels of alcohol consumption in the two-bottle choice paradigm. The Alko Alcohol (AA) and Alko Nonalcoholic (ANA) rat lines were generated in Finland and the sP rat was generated in Sardenia. Two rat lines, alcohol preferring (P) and high alcohol drinking (HAD), were generated in the United States at Indiana University. Both the AA, ANA, sP, P and HAD lines were the result of a bi-directional selection from heterogenous rat stock (Wistar) (Eriksson 1968; Li, Lumeng et al. 1993; Colombo, Agabio et al. 1997). There is much evidence to suggest that these lines of selectively bred

rats may capture multiple aspects of alcohol addiction and further genetic dissection of alcohol related traits in these model system should advance our understanding of the disease in humans (Sinclair, Le et al. 1989; Agabio, Cortis et al. 1996; Lobina, Agabio et al. 1997; Moller, Wiklund et al. 1997; Devoto, Colombo et al. 1998; Bell, Rodd et al. 2006). A comparison of the shared behavioral phenotypes and genotypes in each line selectively bred for high two-bottle choice alcohol consumption could lead to identification of the genetic determinants of each behavior. Bidirectional selection for two-bottle choice alcohol consumption has also led to the development of selectively bred mouse lines. High and low alcohol preference mice (HAP and LAP respectively) originated from an outbred progenitor strain (HS/Ibg) and show a large difference in alcohol consumption and no difference in alcohol elimination or taste discrimination (Grahame, Li et al. 1999). Short-term bidirectional selective breeding from an F2 intercross of two inbred strains of mice with large differences in alcohol consumption (C57BL/6J and DBA/2J) has also been exploited (STS lines) to enhance identification of genes involved in alcohol consumption (Belknap, Richards et al. 1997; Mulligan, Ponomarev et al. 2006).

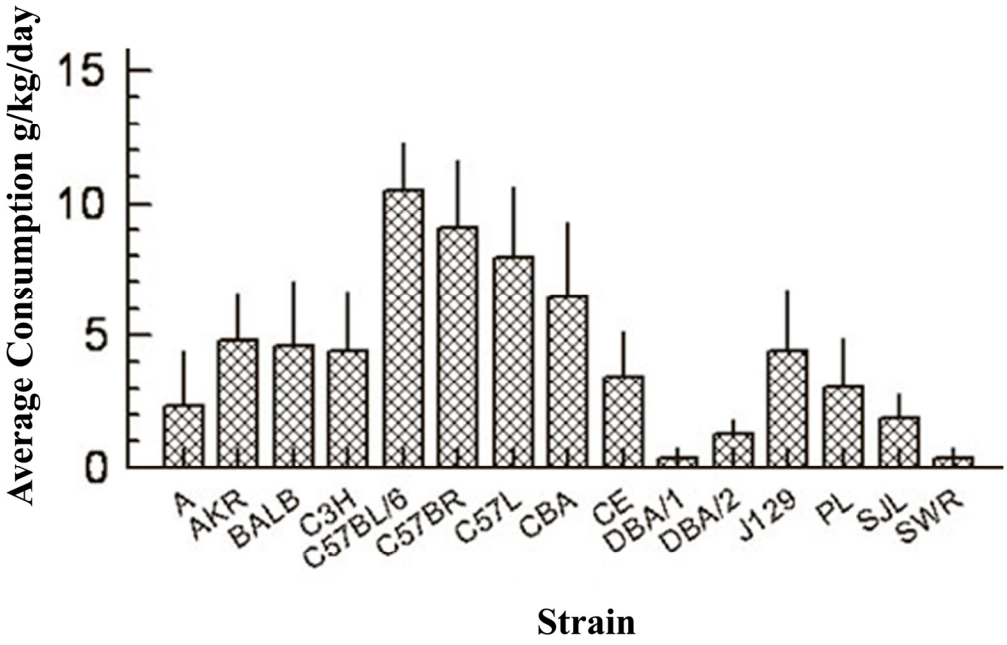
Inbred Strains

Inbred strains of mice are a powerful rodent genetic model available for the elucidation of alcohol related traits. Not only are inbred strains of mice virtually homozygous at all alleles, each inbred strain represents a genetic stock that, with proper quality control measures, is stable over time. This genetic homogeneity has been harnessed to investigate the link between phenotype and genotype and to manipulate specific genes. Essentially, the key to the creation of inbred lines is repeated sibling mating; after about 40 generations of brother and sister mating, nearly all alleles will be homozygous. Many inbred strains are widely available for use in experimental settings

and most have been extensively characterized genetically or genealogically (Beck, Lloyd et al. 2000). The voluntary alcohol consumption in many inbred strains of mice has been well characterized (figure 1.4) and there is a wide range in the levels of alcohol consumption between strains (Belknap, Crabbe et al. 1993). One inbred strain, C57BL/6, consistently consumes more alcohol than any other inbred strain while other inbred strains, such as DBA/2, are stout alcohol non-drinkers. Because of the divergence in alcohol consumption between C57BL/6 (B6) and DBA/2 (D2), both inbred mouse strains have been used as model systems to elucidate the genetic components of alcohol consumption. One technique that has provided many candidate genes for high alcohol consumption in mice is quantitative trait loci (QTL) analysis.

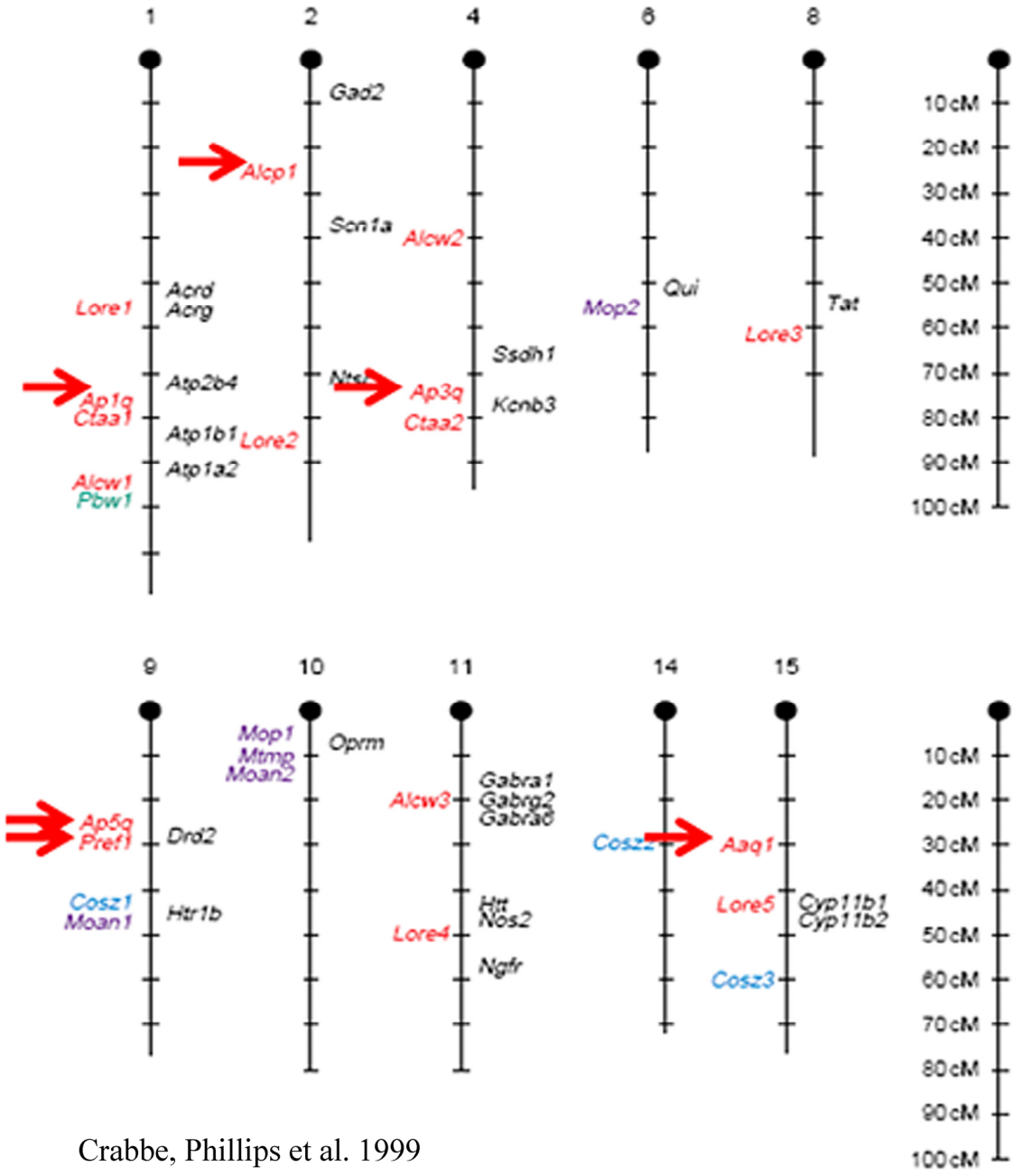
QTL analysis can identify regions of the genome that contain genes associated with specific quantitative phenotypes. Recombinant inbred (BxD RI) strains are created by crossing the B6 and D2 inbred strains. The resulting B6.D2 F1 hybrids are allowed to breed and the BxD RI strains are generated from repeated inbreeding of the F2 population. Each BxD RI strain contains a mixture of parental genes, however, the alleles inherited from each parental strain have essentially been fixed to homozygousity. Therefore, the allelic variation in each BxD RI strain can be correlated with phenotypic variation by QTL analysis. Figure 1.5 provides a summary of the known QTL regions for alcohol preference and consumption between the high alcohol consuming B6 strain and the low alcohol consuming D2 strain. Much like human linkage and association studies, QTL analysis can define broad regions of the chromosome that contain genes related to alcohol consumption in mice. In mice, however, QTL regions can be narrowed in order to identify candidate genes contributing to the phenotype of interest using congenic mice. Congenic strains of mice can be created by backcrossing mice that contain homozygous alleles from one parental strain in a discreet chromosomal region to

the other parental strain. Repeated backcrossing will lead to the creation of mice that contain a small region that contains homozygous alleles from one parent strain while the remaining genome contains relatively homozygous alleles from the other parental strain. Using this technique in BxD RI strains, several regions corresponding to alcohol QTL have been captured that contained B6 alleles within the QTL region and D2 alleles at all other loci in the genome. The reverse case, D2 alleles within the QTL region and B6 alleles at all other locations, can also be generated in this manner. B6xD2 congenic strains can be assayed for alcohol related phenotypes and the QTL interval can be reduced making candidate gene identification more likely. Many congenic strains have been generated, including strains congenic for QTL regions on chromosome 2 and 9 that contain genes involved in alcohol consumption (figure 1.5).



Belknap, Crabbe et al. 1993

Figure 1.4: Voluntary alcohol consumption in inbred mouse strains. C57BL/6 mice consume more alcohol than any other inbred strain. The DBA/2 inbred strain consumes very little alcohol. Consumption is measured in grams of alcohol consumed per kilograms body weight per day (g/kg/day) (Belknap, Crabbe et al. 1993).



Crabbe, Phillips et al. 1999

Figure 1.5: Alcohol preference QTLs. Regions of the mouse genome containing genes associated with alcohol consumption and preference based on QTL analysis in BxD RI strains are indicated by the red arrows. Modified from (Crabbe, Phillips et al. 1999).

Genetically engineered mice

In many ways, rat selected lines offer a more complete model system for alcoholism in terms of behavioral similarities to human alcoholics, however, gene manipulation in mice is far superior to that of rats. The ability to knockout genes or create transgenic animals that have altered expression levels of a gene or express altered genes has provided a wealth of information about the role of candidate genes in behavior and disease. The targeted manipulation of many different genes has been shown to affect the levels of alcohol consumption, however, an analysis of behavioral changes after gene manipulation should always include considerations for the background of the animal in which the manipulation was performed as well as the effect of possible compensatory changes in the system. For brain and behavioral genes, conditional knockouts are often useful for more discreet analysis of gene function. For a detailed survey of alcohol related phenotypes in genetically engineered mice please see (Crabbe, Phillips et al. 2006).

MICROARRAY ANALYSIS

The introduction of microarray technology in 1995, (Schena, Shalon et al. 1995) revolutionized molecular biology and led to the development of new techniques for the global measurement of mRNA, miRNA, genomic DNA, proteins and many other biological molecules. Global analysis of gene expression utilizes the availability of sequenced genomes to create arrays that contain thousands of spots of DNA that represent all or most of the genes of the organism in question. Most microarray platforms consist of arrays that contain either cDNA or oligonucleotide spots. Oligonucleotide arrays are composed of spots containing small oligonucleotides, ~25 nucleotides, designed exclusively from sequence information that can be printed or synthesized directly onto the array (Lipshutz, Fodor et al. 1999). Using a match, mismatch system, oligonucleotide arrays can improve the signal to noise ratio in experimental data but because multiple oligonucleotide probes are used to represent a gene product, space becomes a limiting issue. cDNA arrays consist of longer expressed sequences that represent both known genes and expressed sequences tags (ESTs) of unknown origin. cDNA arrays are relatively inexpensive and more genes can be represented on a single array but the quality control is not as advanced as for oligonucleotide arrays and more steps are involved in cDNA fabrication (figure 1.6). In both cases, the measurement of gene expression involves the collection of mRNA from the samples of interest which are then reverse transcribed into cDNA and labeled with a fluorescent dye (usually cyanine-3, Cy-3, which appears green or cyanine 5, Cy-5, which appears red). Labeled cDNA is hybridized to the microarray and the level of hybridization to each spot is indicated by the level of fluorescence which is detected and quantitated by a laser scanner. Most cDNA arrays compare the hybridization of two samples which yields variations of green,

red and yellow spots depending on the level of binding of each sample to a single position in the array (figure 1.6).

A wealth of microarray data sets have been completed and nearly 4,000 are available for public use from the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo) (Barrett, Troup et al. 2007). Many brain expression analyses have been completed for genetic animal models that were exposed to different alcohol treatments (Worst and Vrana 2005) and for post-mortem brain samples of human alcoholics and controls (Lewohl, Wang et al. 2000; Mayfield, Lewohl et al. 2002; Sokolov, Jiang et al. 2003; Iwamoto, Bundo et al. 2004; Flatscher-Bader and Wilce 2006; Liu, Lewohl et al. 2006). For animal models that have been exposed to alcohol, the different levels of alcohol and the different routes of administration can lead to inconsistency between microarray results. Also, because of the promiscuity of alcohol's actions on the nervous system, alcohol can change the level of many transcripts and this response is often small. These issues might complicate large scale comparisons between microarray datasets for which alcohol was consumed in different quantities. Human microarray studies are very useful in measuring the consequences of long term alcohol abuse on the brain but they were not designed to investigate the role of genetics in the development of alcohol addiction. Microarray analysis between genetic mouse models that differ in voluntary alcohol consumption should elucidate candidate genes that play a role in this behavior. The inclusion of alcohol naïve animals in the gene expression analysis allows for future comparisons across microarray datasets for genetic commonalities between models. The research presented in the next 3 chapters involves the use of microarray analyses to compare brain gene expression between alcohol naïve genetic mouse models that differ in the level of voluntary alcohol consumption. The long term goal of this research is to identify genes that contribute to high voluntary alcohol

intake in mice and, ultimately, to identify candidate genes for alcoholism susceptibility in humans. Portions of the text and some of the figures and tables in chapter 2 were published previously in an article in the *Proceedings of the National Academy of Sciences* in April 2006 (Mulligan, Ponomarev et al. 2006) and are reprinted with the copyright permission policy of The National Academy of Sciences.

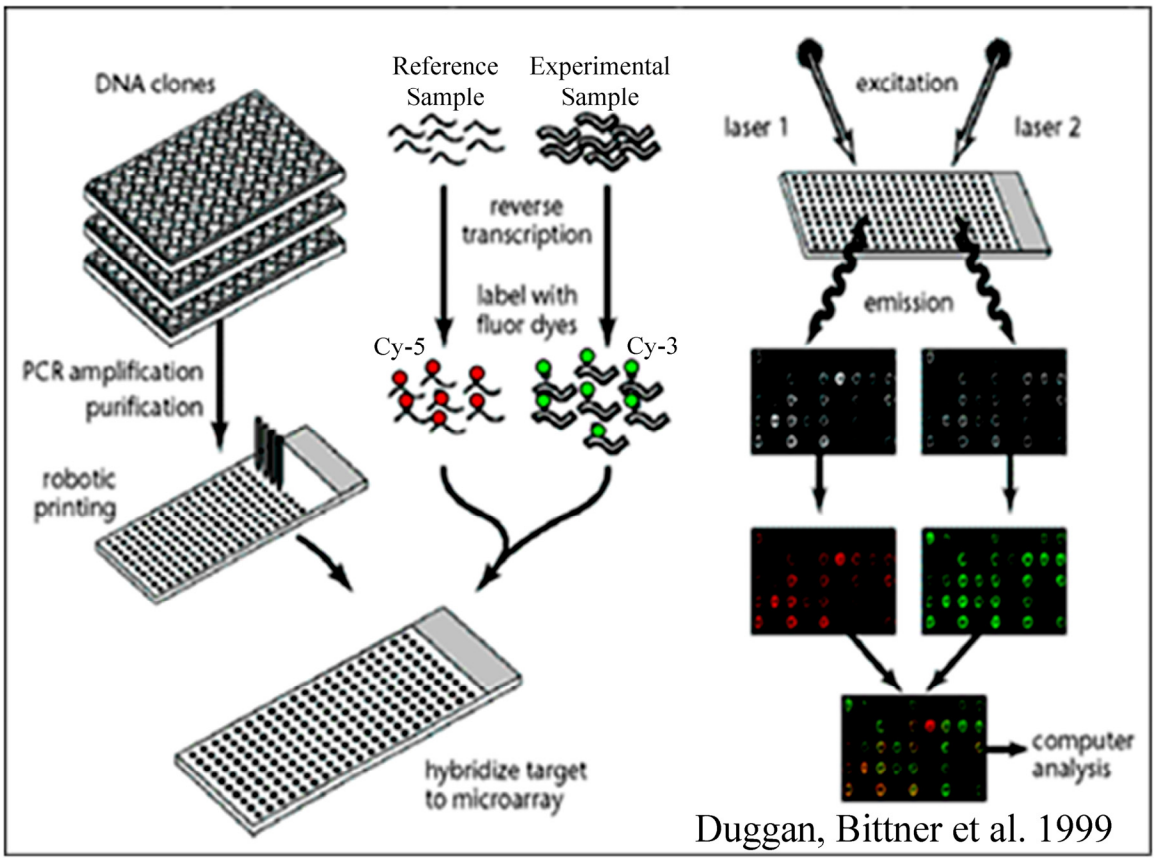


Figure 1.6: Microarray technique. Overview of the cDNA microarray process from microarray printing to hybridization and visualization (Duggan, Bittner et al. 1999).

Chapter 2: Microarray Meta-analysis on Mouse Genetic Models with High and Low Levels of Alcohol Consumption

INTRODUCTION

An understanding of the molecular mechanisms underlying the genetic propensity toward excessive alcohol consumption is crucial for the development of new treatments for alcoholism. Animal models for alcohol-related traits provide an important opportunity to explore mechanisms responsible for different aspects of the uniquely human disease (Crabbe 2002). In particular, selected lines and inbred strains of mice with a divergence in voluntary alcohol drinking represent valuable tools to dissect the genetic components of alcoholism (McClearn 1970; Phillips, Terdal et al. 1990; Grahame, Li et al. 1999; Phillips, Belknap et al. 2002). However, each model has advantages and disadvantages. Selected lines exploit homozygosity for large-effect and some small-effect allelic variants that contribute to the phenotype of divergent alcohol preference and consumption, while the remaining genome remains relatively polymorphic. Each individual selection line is likely to capture a different complement of genes that play a role in the phenotype of interest, even in replicate lines. However, as the selection continues, fixation of genes unrelated to the selected phenotype occurs due to genetic drift. In contrast to selected lines, inbred strains are homogeneous at all alleles. Genetic (allelic) differences between strains lead to observed phenotypic differences, however, any two strains will differ for numerous phenotypes (including gene expression), and identification of the specific genes for a given trait is difficult.

The purpose of this study is to use a meta-analytic approach to identify candidate genes that contribute to the genetic propensity to drink by combining whole-brain gene expression databases of different genetic mouse model systems with divergent voluntary alcohol consumption. In addition to detecting large effect genes, this approach should

allow for the identification of small- or moderate-effect genes. The genetic mouse models used in this study consisted of 3 different sets of oppositely-selected lines bred for high and low alcohol drinking, 5 inbred strains known to differ in voluntary alcohol consumption, and 1 hybrid strain recently shown to have the highest voluntary alcohol consumption of any known mouse genotype (Blednov, Metten et al. 2005) (table 2.1). Global microarray measurement of brain gene expression in each of the above mouse models allowed for the detection of transcripts consistently changed in models with high or low alcohol intake. It is important to note that the mice used for array analysis were not exposed to alcohol; thus, we are defining the transcriptional signatures of genetic predisposition to high and low alcohol consumption. Expression analysis of an additional mouse line congenic for a section of chromosome 9, which contains genes that regulate alcohol consumption (Belknap and Atkins 2001), allowed for the identification of candidate qualitative trait genes (QTGs) for this quantitative trait locus (QTL).

Analysis across microarray datasets for common changes in expression provides more statistical power to detect small and reliable changes than analysis of any one or two of the data sets. Meta-analyses of diverse gene expression datasets has been used previously to successfully uncover genes related to carcinogenic phenotypes (Rhodes, Barrette et al. 2002), but a similar approach has not been previously employed for a neurobehavioral trait (for a review of meta-analyses of microarray data, see Moreau, Aerts et al. 2003). In this study, we employ a meta-analysis of expression data from different mouse genetic models of high and low levels of alcohol consumption to define functional pathways and identify individual genes that may determine a predisposition for high alcohol intake.

Data set	Study	Platform	Strain/line	Sex	*N*
1	STS (B6.D2)	Affymetrix	High STS	F/M	10
			Low STS	F/M	10
2	HAP/LAP1	Affymetrix	HAP1	M	6
			LAP1	M	6
3	HAP/LAP2	Affymetrix	HAP2	M	6
			LAP2	M	6
4	Inbred strains	Affymetrix	B6	M	6
			BalbC	M	6
			D2	M	6
			LP/NJ	M	6
5	Inbred strains	Custom cDNA	B6	F	5
			FVB/NJ	F	5
			FVB.B6 hybrid	F	5
6	B6.D2 congenic 9	Custom cDNA	B6	F/M	12
			B6.D2 Congenic 9	F/M	12

Table 2.1: Mouse models and microarrays used. Three selected lines (data sets 1-3) and the 6 isogenic strains (data sets 4 and 5) were used to form four groups for the meta-analysis. The congenic line (data set 6) was used as a filter to select genes from the meta-analysis. STS, short-term selection; F, female; M, male. Data set 1 was contributed by R.J.Hitzemann, J.C.Crabbe, and J.K.Belknap. Data sets 2 and 3 were contributed by B.Tabakoff. Data sets 5 and 6 were contributed by the University of Texas (M.K.Mulligan and S.E.Bergeson). Previously published in Mulligan, Ponomarev et al. 2006.

RESULTS AND DISCUSSION

Use of Cohen's *d* as an effect size measure allows for the comparison of microarray experiments from diverse sources

Five experimental microarray (Affymetrix and custom cDNA) datasets of mice genetically divergent in measures of voluntary alcohol consumption were analyzed for differential gene expression using the Cohen's *d* statistic (Rosenthal 1994). The initial meta-analysis comprised 13 individual microarray data sets from three groups of selected lines and one group of 6 isogenic strains (table 2.1). A sixth microarray dataset consisting of B6.D2 congenic 9 animals is used later as a filter for the initial meta-analysis. Two advantages of the Cohen's *d* approach are important. First, data from different platforms and different laboratories can be combined without the use of normalization. Second, small differences, which are consistent in the direction of change can be detected, even though these changes would not be detected in any single dataset.

After calculating Cohen's *d* for each data set (figure 2.1A), a pairwise comparison of common changes was applied to test the compatibility between the four datasets and evaluate the use of Cohen's *d* to normalize expression data from separate platforms and different laboratories. The number of transcripts regulated in the same direction between each pair of datasets is significantly greater than chance for five out of six pairs. When all four datasets are compared, the number of transcripts regulated in the same direction is nearly 2-fold over the chance level (χ^2, $p<0.00001$). The lower similarity seen between selected lines and inbred strains, compared to the pairs of selected lines, likely reflect differences in genetic background; an enrichment of alleles linked to alcohol preference in selected lines and a relative randomness of such alleles in inbred strains.

An overlap in regulation between mouse models sharing the same alcohol phenotype is readily detectable despite the use of microarray datasets from different animal models and platforms. Similarly, the distribution of *z*-test *p*-values for the meta-analysis of the four data sets is skewed to the left indicating a high number of significantly co-expressed genes detected by the meta-analysis (figure 2.1B). The shape of the histogram also suggests a low incidence of false positives consistent with the calculated *Q* values. Thus, the use of Cohen's *d* in our meta-analysis allows for the detection of transcripts significantly divergent between genetically distinct high and low alcohol consuming mouse models.

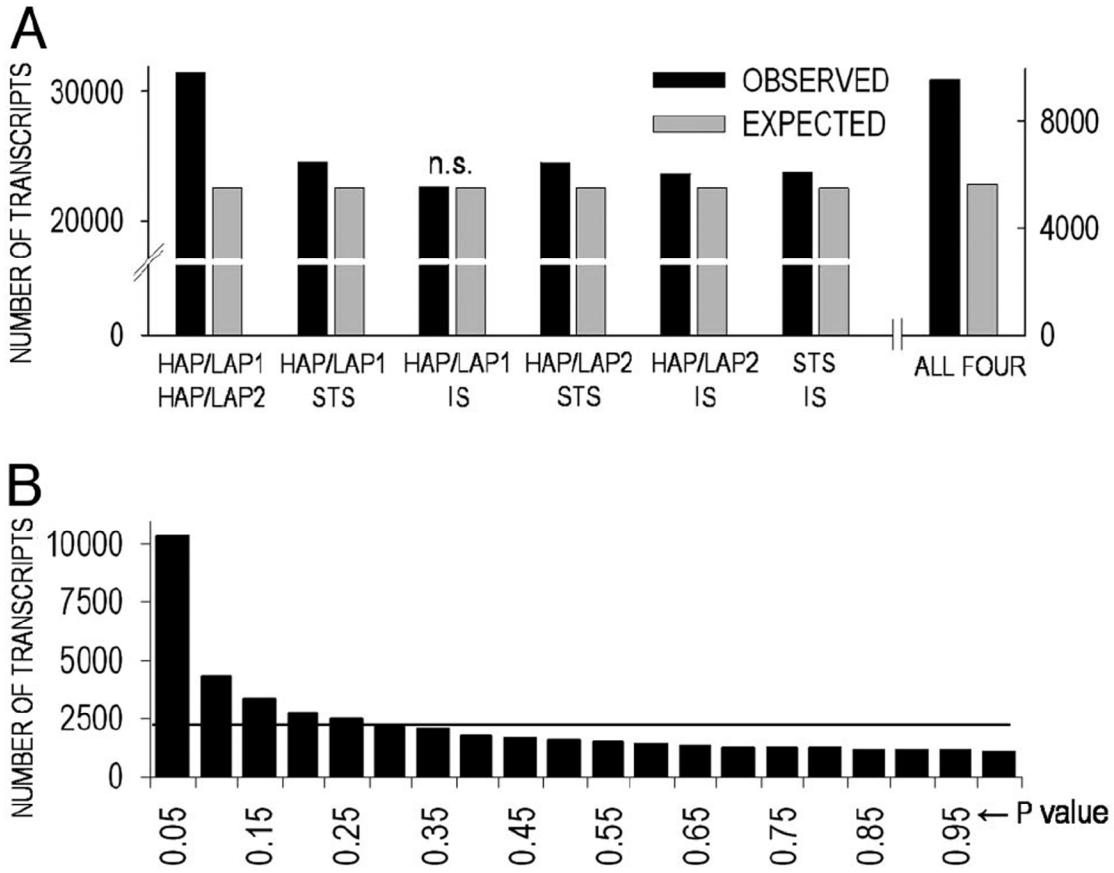


Figure 2.1: Comparison of microarray data sets. (A) The number of transcripts regulated in the same direction between any two (left y axis) or all four (right y axis) data sets ($P<0.0001$, χ^2). n.s., not significant. (B) Frequency distribution of z test P values (x axis) is shown. The solid line represents theoretical chance distribution. IS, inbred strain. Previously published in Mulligan, Ponomarev et al. 2006.

Meta-analysis identifies candidate genes for high and low levels of alcohol consumption

Use of Cohen's *d* values between the high and low alcohol drinking mouse strains and lines across datasets resulted in the identification of 5,182 significant ($|d| < 0.5$ and $Q < 0.05$) differentially expressed transcripts (supplemental table of all significant genes available on the PNAS website at http://www.pnas.org), representing 3,800 unique genes. The meta-analysis results are visualized as a remarkable consistency in gene expression changes across the four microarray datasets when displayed as a pseudo-color raster of the highly significant ($Q < 0.01$) transcripts (figure 2.2A). Red and green indicate transcripts over and under-expressed in high alcohol consuming animals, respectively. In general, more transcripts appear to be up-regulated in the high consuming models relative to low consuming models.

The sheer number of significant differences suggests that numerous pathways in the brain may be altered between high and low alcohol consuming mouse models. Indeed, the seventy-five genes from figure 2.2 with the largest absolute effect size and lowest q-value (figure 2.2B) fall into the broad categories of cellular homeostasis and neuronal function. Several genes of interest include the up-regulated (expressed higher in high drinking compared to low drinking mice) genes *B2m*, *Man2b1*, *Scn4b*, *Mapre1*, *Prkce* and *Sst* which function in immunity/cellular defense, glycosylation, ion channel activity, microtubule binding/dynamics, intracellular signaling cascades and neuronal signaling, respectively. *B2m*, *Scn4b* and *Prkce* are genes of special interest because they have previously been identified as important for neuronal functions. *B2m* (β2 microglobulin) may play a role in neuronal plasticity as *B2m* knockout mice exhibit abnormal synaptic connections as well as altered patterns of long-term potentiation and long-term depression (Huh, Boulanger et al. 2000). *Scn4b* (sodium channel β4) is an

auxiliary subunit of voltage-gated sodium channels which can alter channel function as well as the interaction between the ion channel and other proteins (Yu, Westenbroek et al. 2003). In cerebellar Purkinje cells, *Scn4b* may be required for high-frequency firing of voltage-gated sodium channels (Grieco, Malhotra et al. 2005). A role for *Prkce* (protein kinase C-ε) in ethanol consumption is supported by the observation that *Prkce* knockout mice consume less alcohol when compared to controls in a two-bottle choice drinking paradigm and introduction of conditional *Prkce* expression in these mice was sufficient to restore alcohol drinking levels (Hodge, Mehmert et al. 1999). *Gnb1*, *Hmgn2* and *Hyou* are significantly down-regulated in alcohol-preferring mice compared to non-preferring mice. *Gnb1*, *Hmgn2*, and *Hyou* function in GTPase activity/signal transduction, DNA binding/packaging and the cellular response to stress, respectively. The overall results of the meta-analysis show the complexity of the alcohol drinking phenotype as indicated by the surprisingly large number of consistent, small increases or decreases in gene expression between the mouse models.

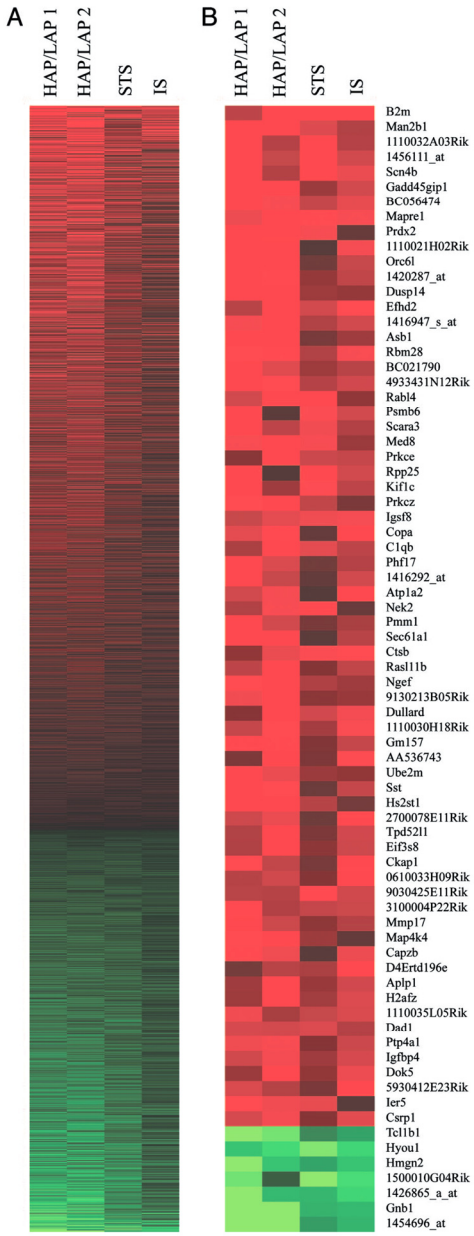


Figure 2.2: Visual representation of microarray data. The criterion for inclusion was $Q < 0.01$. A positive effect size (red) represents high expression in preferring mouse models and a negative effect size (green) represents lower expression in preferring mice. Black indicates an effect size close to zero. (A) Transcripts (2,697) sorted from high to low effect size. (B) The top 75 transcripts based on highest and lowest effect size ($Q<0.05$, $|d| > 1.94$). STS, short term selection; IS, inbred strain. Previously published in Mulligan, Ponomarev et al. 2006.

In silico overrepresentation analyses reveal potential significant regulation by transcription factors and functional pathways

Nearly 10% of the mouse genome was detected as being divergent between the high and low alcohol-preferring mouse models. It is possible that linkage disequilibrium could account for the surprisingly large number of consistent changes detected by the meta-analysis. Linkage disequilibrium is the tendency of closely linked genes to be co-expressed. Another explanation could be that some of the risk-conferring QTG are transcriptional regulators. To test the later hypothesis, the oPOSSUM database (http://www.cisreg.ca) (Ho Sui, Mortimer et al. 2005) was used to detect overrepresented transcription factor binding sites (TFBSs). oPOSSUM identifies transcription factor binding sites (TFBSs) in conserved (orthologous between human and mouse) non-coding promoter regions located within 0-2000 bp upstream of a gene's transcription start site. Current *in silico* methods for promoter analysis are still evolving and it is not yet possible to complete an exhaustive analysis of all possible TFBS in the mammalian genome. Because it is likely that a large number of potential TFBS will be overlooked due to the previously discussed shortcomings involved with promoter analysis, a very conservative approach was taken in order to minimize the number or false positives and maximize the likelihood that the few TFBS that could be detected would be statistically more likely to be accurate. The analysis provided by the oPOSSUM database is very amenable to this goal and is very conservative with strict requirements; simulation studies have suggested that it detects few false positives (Ho Sui, Mortimer et al. 2005).

Significant transcripts from the meta-analysis were separated into two distinct lists containing transcripts up- and down-regulated in preferring mice relative to non-preferring mice. Although there were more up-regulated than down-regulated transcripts evaluated by oPOSSUM, only 4 transcription factor binding sites were found

to be significantly overrepresented in up-regulated genes, whereas eighteen transcription factor binding sites were found to be significantly overrepresented in down-regulated genes (table 2.2). HRF-2, SAP-1, Bsap and STAF binding sites were found to be overrepresented in up-regulated genes. Several of the up-regulated genes had multiple TFBSs in their upstream promoter regions suggesting the involvement of a common regulatory network or control pathway for alcohol preference. It should be noted that some TFBSs (*e.g.* HFH-2 and HFH-3-1; NRF-2 SAP) have similar consensus sequences, which could lead to similar rankings in oPOSSUM. The TFBS for Staf, (selenocysteine RNA gene transcription activating factor) is recognized by *Zfp143* and was identified as overrepresented in the up-regulated genes (for high alcohol drinking). The STAF TFBS was present in the upstream (2000 bp) promoter regions of 64 genes. Importantly, *Zfp143* expression was also found to be significantly up-regulated across the high drinking models. The TFBS for the fork-head box transcription factor *Foxa2* (HNF-3 β, hepatocyte nuclear factor 3β) was detected as overrepresented across the down-regulated gene group and present in 146 genes. Again, *Foxa2*, itself, was found to be significantly down-regulated in the meta-analysis. Fork-head transcription factors are known to regulate cell differentiation and tissue development, are involved in immune and hormone responses, and have recently been shown to regulate neuronal function and overall lifespan and other biological processes too numerous to mention here (Coffer and Burgering 2004; Fukunaga, Ishigami et al. 2005; Grieco, Malhotra et al. 2005; Laganiere, Deblois et al. 2005; Morris 2005; Phillips, Broadbent et al. 2005). In both the case of *Foxa2* and *Zfp143*, the overrepresented TF has the same pattern of expression as its target genes and is an example of how a change in the expression of a particular transcription factor could potentially account for changes in the levels of many genes. It is easy to imagine, as evidenced by the large number of gene targets for both Zfp143 and Foxa2,

how small changes in the levels of a transcription factor could cause profound differences in the brain transcriptome and may account for the large number of gene expression changes (3,800 genes) detected in this meta-analysis. An understanding of how changes in the expression level and/or activity of transcriptional regulators affects the predisposition towards alcohol consumption phenotypes warrants further investigation.

In order to better understand possible global interactions of the divergent genes, over-representation analyses for function were completed. Such analyses apply statistical methods to estimate whether some biological function/pathway is represented in a given dataset more than expected by chance. Using the WebGestalt search engine (http://genereg.ornl.gov/webgestalt/), which queries different databases including KEGG, BioCarta and GO (gene ontology) (Zhang, Kirov et al. 2005), functional group analysis revealed that kinase and signaling pathways were overrepresented in genes divergent between alcohol-preferring and non-preferring genotypes. The result supports previous suggestions that the MAPKinase, NF-KB and IL6 signaling pathways (and CREBBP) are sensitive to alcohol and alcohol withdrawal (Jeong, Hong et al. 2005). Known functional interactions (summarized in figure 2.3) show that gene transcription pathways predominant in the overrepresented categories and apoptosis/anti-apoptosis, neurodegeneration, and neuroplasticity pathways were also functionally over-represented (for a complete list see tables 2.3 and 2.4). *In silico* functional analyses are powerful tools for hypothesis generation, especially for complex phenotypes such as predisposition toward alcohol drinking. It should now be feasible, for example, to test the effect of the bioactive peptides shown in figure 2.3 for their impact on alcohol consumption.

Up-regulated in Preferring					
TFBS	**TF**	**BH**	**TH**	**p**	***Z***
NRF-2	**Gabpa**	1245	237	**0.005**	*13.39*
SAP-1	Elk4	1110	210	**0.011**	*12.55*
BSAP	**Pax5**	235	52	**0.024**	*11.67*
STAF	**Plagl2**, *Zxda*, **Zfp143**	313	64	**0.047**	*13.29*

Down-regulated in Preferring					
TFBS	**TF**	**BH**	**TH**	**p**	***Z***
HFH-1	Foxq1	696	94	**3.1E-5**	*12.75*
Nkx	Nkx2-5	3203	325	**8.7E-5**	*8.85*
S8	*Prx2*	2759	276	**0.002**	*5.15*
SRY	Sry	2401	242	**0.003**	*2.31*
FREAC-4	Foxd1	1298	141	**0.003**	3.73
Gfi	Gfi1	2247	227	**0.004**	*0.17*
Thing1-E47	**Hand1**, Tcf3, **Tcfe2a**	2237	226	**0.004**	*1.56*
HNF-3beta	**Foxa2**	1373	146	**0.006**	*1.67*
HFH-3	Foxi1	1390	147	**0.007**	*4.23*
SOX-5	**Sox5**	2530	250	**0.008**	*0.87*
HFH-2	Foxd3	1459	152	**0.010**	*3.03*
ROR alpha 1	Rora	978	104	**0.022**	*2.70*
c-MYB 1	Myb	2358	229	**0.026**	*-2.44*
COUP-TF	Ear3, Nr2f1	446	52	**0.026**	*5.80*
AML-1	Runx1	1928	189	**0.035**	*1.93*
SPI-B	Spib	4255	391	**0.039**	*0.61*
SOX17	**Sox17**	2600	247	**0.048**	*1.56*
HNF-1	Tcf1	254	31	**0.048**	*7.90*

TFBS = Transcription Factor Binding Site, TF = Transcription Factor, BH = Background Hits from all genes represented in meta analysis, TH = Target Hits from gene list, p = Fisher Exact Probability from oPOSSUM and Z = Z-score from oPOSSUM. **Red** = significantly ($q < 0.05$) up-regulated in preferring, **green** = significantly ($q < 0.05$) down-regulated in preferring, *italics* = not represented in meta analysis, plain = not significant.

Table 2.2: Overrepresented transcription factor binding sites. oPOSSUM was used to find transcription factor binding sites that were overrepresented for genes up-regulated in preferring mice and down-regulated in preferring mice. Genes from the meta-analysis that had $q<0.05$ and an average $|d|<0.5$ were selected for transcription factor binding site overrepresentation analysis. This gene list was separated into genes up-regulated in preferring mice (2,388 with 1,011 analyzed in oPOSSUM) and genes down-regulated in preferring mice (1,580 with 546 analyzed in oPOSSUM). Criterion for significant overrepresentation was $p<0.05$. Interestingly, fewer binding sites are found to be significantly overrepresented in the genes up-regulated in preferring animals.

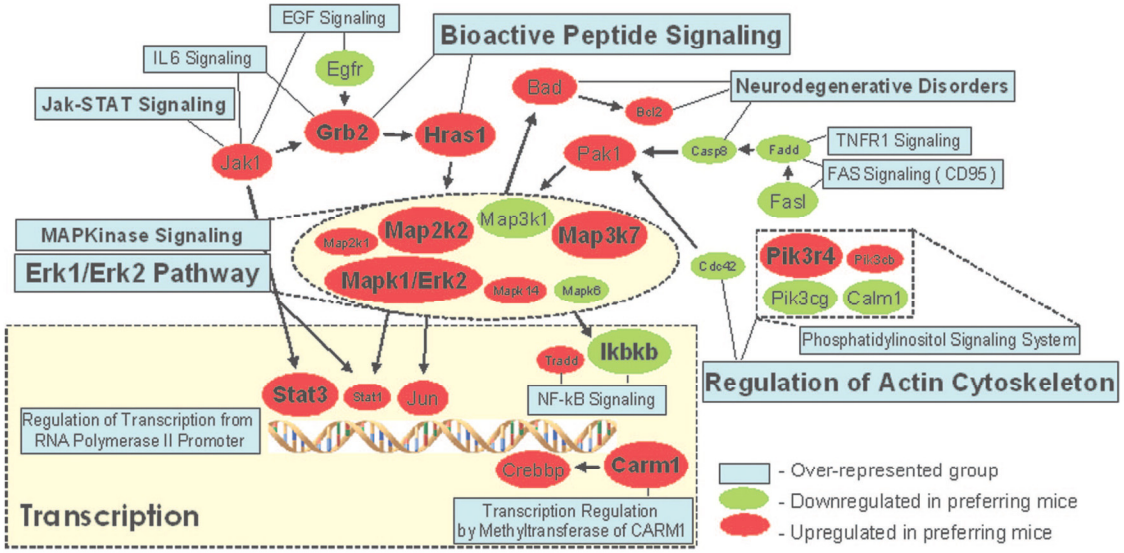


Figure 2.3: Functional pathways altered between high and low alcohol consuming mouse models. Genes present in at least three overrepresented functional groups/pathways from BioCarta, KEGG and Gene Ontology are shown. Larger font sizes represent either smaller P values for functional groups/pathways or larger effect size for individual genes. (p<0.001, Q<0.01 and $|d|$>0.5 for all genes). Lines connect gene symbols with relevant functional groups. Arrows indicate pathway connections. Previously published in Mulligan, Ponomarev et al. 2006.

Biocarta pathways	Gene number	Entrez Gene IDs	P-value
		13649 14688 14699 14784 15461 16412 17346 19279 20848	
Erk1/Erk2 Mapk Signaling pathway	14	26395 26396 26413 51792 73086	7.87E-05
Map Kinase Inactivation of SMRT Corepressor	7	13649 19401 20181 20602 26395 26401 26416	0.00032
		12313 14061 14672 14688 14699 14784 15461 17762 19229	
Bioactive Peptide Induced Signaling Pathway	14	20846 26395 26396 26413 26416	0.00061
		12313 12929 14784 15461 16476 18479 19229 26395 26396	
Links between Pyk2 and Map Kinases	12	26401 26413 26416	0.00061
NFAT and Hypertrophy of the heart (Transcription in the		12313 12914 13019 14061 14173 14463 15110 15200 15461	
broken heart)	15	16878 17260 26395 26413 26416 72508	0.0018
		12015 12028 12043 12370 14082 20597 26395 26401 26413	
Ceramide Signaling Pathway	10	71609	0.00266
		12313 14293 14688 14699 15461 17969 18479 26395 26396	
fMLP induced chemokine gene expression in HMC-1 cells	12	26401 26413 26416	0.00355
Role of ß-arrestins in the activation and targeting of MAP			
kinases	7	110355 14688 14699 16485 26395 26396 26413	0.00398
ER–associated degradation (ERAD) Pathway	7	14376 17159 21402 22196 230904 53421 56438	0.00398
Roles of ß-arrestin-dependent Recruitment of Src Kinases			
in GPCR Signaling	8	110355 14688 14699 15461 16485 26395 26396 26413	0.00433
Ras-Independent pathway in NK cell-mediated cytotoxicity	8	12010 16412 16642 16797 18479 19229 20963 26395	0.00433
		12929 14784 15234 15461 16412 16476 18479 19229 19247	
Signaling of Hepatocyte Growth Factor Receptor	13	20848 26395 26396 26413	0.00451
IL22 Soluble Receptor Signaling Pathway	5	12702 16154 16451 20846 20848	0.00475
Angiotensin II mediated activation of JNK Pathway via		12313 13649 14784 15461 16476 18479 19229 26395 26396	
Pyk2 dependent signaling	11	26401 26413	0.00569
Role of ERBB2 in Signal Transduction and Oncology	9	13649 13867 13982 14784 15461 20848 26395 26413 59035	0.0068
		12914 19014 20181 20185 20588 22084 22385 59035 66354	
Control of Gene Expression by Vitamin D Receptor	10	93760	0.00914
		12015 12043 12048 12702 12929 14103 14784 15461 16184	
IL-2 Receptor Beta Chain in T cell Activation	13	16451 20963 268373 72508	0.01115
PTEN dependent cell cycle arrest and apoptosis	8	12576 14103 14784 16202 16412 18604 26413 56484	0.01115
		14784 15461 16150 16476 17346 18479 20112 20846 23938	
		26395 26396 26400 26401 26409 26413 26414 26416 269881	
MAPKinase Signaling Pathway	23	50772 71609 72508 73086 76089	0.01391
Cyclins and Cell Cycle Regulation	9	12428 12443 12444 12447 12530 12534 12575 12576 12578	0.01472
WNT Signaling Pathway	9	12234 12443 12914 13016 14825 18099 19015 26409 51792	0.01472
		12370 14082 14084 14103 16476 16907 18479 19090 20740	
FAS signaling pathway (CD95)	11	26401 26409	0.01521
		12313 14126 14127 14784 15461 16476 20963 26395 26400	
Fc Epsilon Receptor I Signaling in Mast Cells	11	26401 26413	0.01521
IL 6 signaling pathway	8	12608 14784 15461 16451 16476 19247 20848 26395	0.0166
Mechanism of Acetaminophen Activity and Toxicity	4	12355 13106 19224 19225	0.01718
		12043 13649 14103 15461 16150 16476 23871 26395 26400	
Keratinocyte Differentiation	13	26401 26413 26416 51792	0.01867
NF-kB Signaling Pathway	8	14082 16150 16151 16177 16179 26401 26409 71609	0.02372
mTOR Signaling Pathway	8	13669 13684 17346 18604 20104 22084 51792 72508	0.02372
Role of MAL in Rho-Mediated Activation of SRF	8	12540 13367 15461 16885 26395 26396 26401 26413	0.02372
Regulation of BAD phosphorylation	7	12015 12028 12043 12048 12984 16590 26413	0.02712
Role of nicotinic acetylcholine receptors in the regulation			
of apoptosis	6	11443 12015 14103 19229 21752 56484	0.03051
		12370 14082 16476 16907 18479 19090 20740 26401 26409	
TNFR1 Signaling Pathway	10	71609	0.03106
Inhibition of Cellular Proliferation by Gleevec	8	12015 12929 14784 15461 16476 20846 26395 26401	0.03274
		12475 16150 16151 16179 16476 19106 21899 21946 26401	
Toll-Like Receptor Pathway	11	26409 26416	0.03342
Transcription Regulation by Methyltransferase of CARM1	4	12914 19401 20181 59035	0.03375
EGF Signaling Pathway	9	13649 14784 15461 16451 16476 20846 20848 26395 26401	0.03688
Signal transduction through IL1R	9	15977 16150 16177 16179 16181 16476 26401 26409 26416	0.03688
Trefoil Factors Initiate Mucosal Healing	8	12015 13649 14600 14784 15461 16412 21785 26413	0.04382
Sprouty regulation of tyrosine kinase signals	6	13649 14784 15461 24064 26395 26413	0.04452

Table 2.3: Overrepresented Biocarta Pathways from the meta-analysis. All pathways significantly overrepresented at the $p<0.05$ level are shown. Published previously as supporting information in Mulligan, Ponomarev et al. 2006.

KEGG pathways	Gene number	Entrez Gene IDs	P-value
		101314 105855 107589 109880 11567 12475 12540 12632 12669 12927 12929 13367 13649 14061 14166 14173 14178 14182 14673 15461 16402 16403 16411 16412 16419 16885 17179 17532 17698 18479 18591 18596 18645 18718 18719	
Regulation of actin cytoskeleton	64	19684 20130 214230 216963 22349 225724 24 107476 11532 11669 11677 14651 16828 16832 17448 17449	1.04E-05
Pyruvate metabolism	16	18563 18746 18770 235339 60525 66204 68263 104112 11429 12974 15929 17448 17449 18563 269951 66925	3.10E-05
Citrate cycle (TCA cycle)	12	66945 67834 78920	5.13E-05
Glycolysis / Gluconeogenesis	20	11529 11532 11669 11676 12183 13806 14121 14751 15275 16828 16832 18642 18746 18770 235339 56421 58810 60525 66204 68263	0.00018
Pyrimidine metabolism	22	110074 11628 12495 18103 18971 18974 19076 20016 20021 20022 20135 22247 22271 56248 66420 67464 69833 70408 71701 72962 78929 80914	0.00165
MAPK signaling pathway	63	109689 109880 12363 12370 12475 12540 12929 13649 14103 14166 14173 14178 14182 14673 14784 15461 15507 16150 16151 16177 16178 16476 16765 17260 17346 17532 17762 17873 18099 18479 18591 18596 18780 19042 19057 19060 19279 19417 20130 216869 21809 225724	0.00203
Focal adhesion	55	107589 109880 11796 12015 12043 12143 12333 12336 12443 12444 12540 12927 12929 13367 13649 14784 15234 15461 16202 16402 16403 16412 16419 16476 16773 16775 16777 17532 18479 18591 18596 20130 20418 214230 21827 21828 22165 22339 22370 22371 226519 23830	0.00239
Carbon fixation	10	108682 11676 14121 14719 17448 17449 18746 18770 21881 83553	0.00266
Reductive carboxylate cycle (CO2 fixation)	5	11429 17448 17449 269951 60525	0.00475
Tight junction	32	110854 114141 12540 12737 13822 14679 15461 17532 18754 18762 19052 19053 19054 20130 20336 20740 20742 211446 21873 22612 51792 56217 56449 56492 56513 57255 60595 66922 67374 69908 71960 73699	0.00628
Jak-STAT signaling pathway	37	12048 12443 12444 12702 12804 12914 12981 12984 13019 14600 14784 15977 16153 16154 16160 16162 16184 16195 16199 16451 16878 16880 17480 18414 20846 20848 24064 24066 30955 54607 56324 56469 57914 67296 74769 75669 93672	0.00806
Neurodegenerative Disorders	12	12015 12028 12043 12048 12370 12914 13498 14784 17762 19122 19164 57320	0.00973
Adherens junction	21	11431 12540 12914 13649 14182 17126 18099 19246 20411 20583 225724 22612 242687 245880 26409 26413 29875 50772 58235 60595 99167	0.0113
Cell cycle	28	107995 12236 12237 12428 12443 12444 12447 12530 12534 12575 12576 12578 12914 17120 17216 17219 17220 17873 18538 18817 19090 21809 22628 23882 50496 50793 56452 70315	0.01376
Fatty acid metabolism	16	113868 11430 11529 11532 11669 13076 13090 13094 13099 13106 13109 14081 15107 216739 23986 93732	0.01419
RNA polymerase	7	20016 20021 20022 66420 69833 70408 78929	0.01838
One carbon pool by folate	6	107747 107885 108147 14317 14450 20425	0.01972
Galactose metabolism	9	11605 11677 14387 15275 18642 27384 56348 56421 69976 102414 110197 114663 12313 13631 14773 15258 16438 16439 18718 18719 18797 19106 30955 52858 55980 63828	0.02055
Phosphatidylinositol signaling system	20	74769 75669 80796 12313 12370 12411 12914 13191 14784 20185 53323 67300	0.02281
Huntington's disease	10	80796	0.02372
Prion disease	5	12043 16777 16785 19122 226519	0.03325
Amyotrophic lateral sclerosis (ALS)	7	12015 12028 12043 12048 14775 20832 85305	0.0383
Aminoacyl-tRNA biosynthesis	8	105148 107271 20226 234734 23874 27267 353172 85305 11656 12411 12660 20226 20425 26912 27384 353172 56348	0.04382
Glycine, serine and threonine metabolism	10	67092 11529 11532 13123 13190 14719 15233 15445 15468 22173	0.05024
Tyrosine metabolism	11	22178 71974	0.0519

Table 2.4: Overrepresented KEGG Pathways from the meta-analysis. All pathways significantly overrepresented at the $p<0.05$ level are shown. Published previously as supporting information in Mulligan, Ponomarev et al. 2006.

Several genes identified in the meta-analysis are located within QTLs for human alcoholism vulnerability

A two-pronged approach was taken to utilize known QTLs in both human and mouse research. First, the list of ~3800 significant genes was annotated using SOURCE (Diehn, Sherlock et al. 2003) (http://source.stanford.edu) for both mouse and human chromosomal location. Two separate lists were created using Feighner, DSM-III-R and ICD-10 criteria (Foroud, Edenberg et al. 2000): one list for those genes found within boundaries of mouse QTL for preference drinking (Belknap and Atkins 2001) and another for genes within boundaries of alcoholism susceptibility QTL in humans. The lists were used as filters to identify overlap of significant genes from the meta-analysis with QTL between the two species (table 2.5). Thirty-six genes met filter criteria as candidate genes with 11 of the genes (*Gstm1, Gstm2, Gstm5, Il1f5, Il1f6, Il1f8, Il1rn, S100a1, S100a10,* and *S100a13)* belonging to only three gene families: GSTM (glutathione S-transferase activity), S100 class calcium binding, and cytokine/proinflammatory activity (IL1) (table 2.6). Little attention has been given to a role for these genes in alcohol consumption, but GSTM1 alleles are associated with increased alcoholism and alcoholic liver disease (Engracia, Leite et al. 2003). Mice with mutant chemokine receptors suggest a role of inflammatory pathways in alcohol consumption (Blednov, Bergeson et al. 2005). QTL analysis for alcohol preference in rats (Bice, Foroud et al. 1998; Carr, Habegger et al. 2003) show two syntenic QTLs with mouse and the chromosome 2 QTL is syntenic between all three species, including humans. GST8-8, the implicated QTG within the QTL on rat chromosome 8 (Liang, Habegger et al. 2004) is syntenic on mouse chromosome 9. However, Gsta4, the mouse alias for GST8-8, was not significantly divergent in this meta-analysis. Interestingly, several other significantly regulated glutathione-S-transferase genes found on mouse

chromosome 1 and human chromosome 3 were in syntenic QTL regions for both species (table 2.6).

It is important to note that although there is considerable synteny between the genomes of mice and man, the degree of evolutionary overlap between polymorphisms or alleles of genes is probably much lower. Thus, translation of data from mouse models to humans is more likely to apply on a pathway level rather than as exact mutations in specific genes. The specie differences in glutathione-S-transferase may, in fact, be such an example.

Mouse QTL

Source	Chromosome	Location	Reference
BXD, B6.D2 F2	2	14-78 Mb	1
	3	80-139 Mb	1
	4	99-123 Mb	1
	9	14-96 Mb	1

Human QTL

Source	Chromosome	Location	Reference
COGA data	1	73-144 Mb	2
	2	54-127 Mb	2
	3	55-112 Mb	2
	7	47-74 Mb	2
	8	68-91 Mb	2
	22	11-37 Mb	2

Syntenic QTL Regions

Mouse Chromosome	Human Chromosome
2:24.1-24.3 Mb	2q11.1-14.2
3:88-122 Mb	1p22-1q31
4:100-104 Mb	1p22-1q31

1. Belknap, J. K. & Atkins, A. L. (2001) Mamm. Genome **12**, 893–899.

2. Foroud, T., Edenberg, H. J., Goate, A., Rice, J., Flury, L., Koller, D. L., Bierut, L. J., Conneally, P. M., Nurnberger, J. I., Bucholz, K. K., et al. (2000) Alcohol. Clin. Exp. Res. **24**, 933–945.

Table 2.5: Overlapping QTL between rodents and humans. Mouse QTL regions for alcohol preference are shown in the top panel and Human QTL for alcoholism are shown in the middle panel. The bottom panel shows shared QTL regions between mouse and human. References are provided for the QTL information at the bottom of the table. Previously published as supporting information in Mulligan, Ponomarev et al. 2006.

Symbol	Chr	loci	Symbol	Chr	loci
Abcd3	3	122.14	Kcnc4	3	107.82
	1	1p22-p21		1	1p21
Anxa9	3	95.29	Mg29	3	108.59
	1	1q21		1	1p13.3
Arhgef11	3	88.13	Olfm3	3	115.25
	1	1q21		1	1p22
Atp1a1	3	101.51	Palmd	3	117.26
	1	1p21		1	1p22-p21
Bcar3	3	122.8	Pde4b	4	100.6
	1	1p22.1		1	1p31
Cept1	3	106.88	Phtf1	3	103.9
	1	1p13.3		1	1p13
Ctss	3	95.52	Pmvk	3	89.88
	1	1q21		1	1p13-q23
Ddx20	3	106	Ppap2b	4	103.5
	1	1p21.1-p13.2		1	1pter-p22.1
Gba	3	89.63	Prcc	3	88.35
	1	1q21		1	1q21.1
Gnai3	3	108.49	Prpf3	3	95.83
	1	1p13		1	1q21.1
Gpsm2	3	109.06	Rhoc	3	104.74
	1	1p13.3		1	1p13.1
Gstm1	3	108.39	S100a1	3	90.94
	1	1p13.3		1	1q21
Gstm2	3	108.36	S100a10	3	93.77
	1	1p13.3		1	1q21
Gstm5	3	108.28	S100a13	3	90.94
	1	1p13.3		1	1q21
Il1f5	2	24.24	Sort1	3	108.66
	2	2q14		1	1p21.3-p13.1
Il1f6	2	24.18	Tcfl1	3	95.11
	2	2q12-q14.1		1	1q21
Il1f8	2	24.11	Thbs3	3	89.64
	2	2q12-q14.1		1	1q21
Il1rn	2	24.3	Txnip	3	96.56
	2	2q14.2		1	1q21.1

Table 2.6: Candidate genes located within human and mouse overlapping QTLs for alcoholism risk or alcohol preference. Shaded rows indicate mouse genes and rows without shading indicate human gene. Mouse loci are based on physical map location (Mb) and human loci are based on cytological map position. Chr, chromosome.

Use of a B6.D2 congenic 9 data set filter identifies *cis*-regulated genes occupying a strong alcohol preference QTL in mice

QTL mapping and construction of congenic mice based on such maps provides another genetic approach to defining changes in gene expression linked to propensity for high alcohol consumption. Several mouse QTL experiments show that chromosome 9 contains genes that contribute to the genetic propensity toward high alcohol intake (Belknap and Atkins 2001). Therefore, C57BL/6J congenic mice containing a region that captured the DBA/2J alleles between 9-61 cM on chromosome 9 were analyzed for gene expression, and compared against the C57BL/6J background strain. The differentially expressed genes located on chromosome 9 are by default *cis*-regulated and thus are QTG candidates.

In order to identify *cis*-acting candidate genes for the chromosome 9 alcohol preference QTL, the current meta-analysis results were further filtered using the congenic dataset. Significant genes detected by the meta-analysis that passed the congenic filter were mapped on chromosome 9 along with their relative effect size (figure 2.4). These genes include 16 known genes (*Arhgef12, Carm1, Cryab, Cox5a, Dlat, Fxyd6, Limd1, Nicn1, Nmnat3, Pknox2, Rbp1, Sc5d, Scn4b, Tcf12, Vps11, Zfp291)* and four expressed sequence tags (*2810423O19Rik*, *1110032A03Rik*, *5730439E10Rik* and *9030425E11Rik*). Four of these, *Pknox2, Scn4b, 9030425E11Rik* and *1110032A03Rik,* are correlated with alcohol preference in a BxD recombinant inbred population (Phillips, Crabbe et al. 1994). Interestingly, some primary candidate genes also have putative transcriptional activity: *Carm1*, *Pknox2*, *Tcf12*, *Zfp29* and *Limd1*. Although their targets are not yet characterized, each may explain some of the consistently divergent gene expression observed in the meta-analysis. The 20 genes identified with the congenic

filter are potential QTGs underlying alcohol preference and provide exciting avenues of investigation into the complex behavior of alcohol preference.

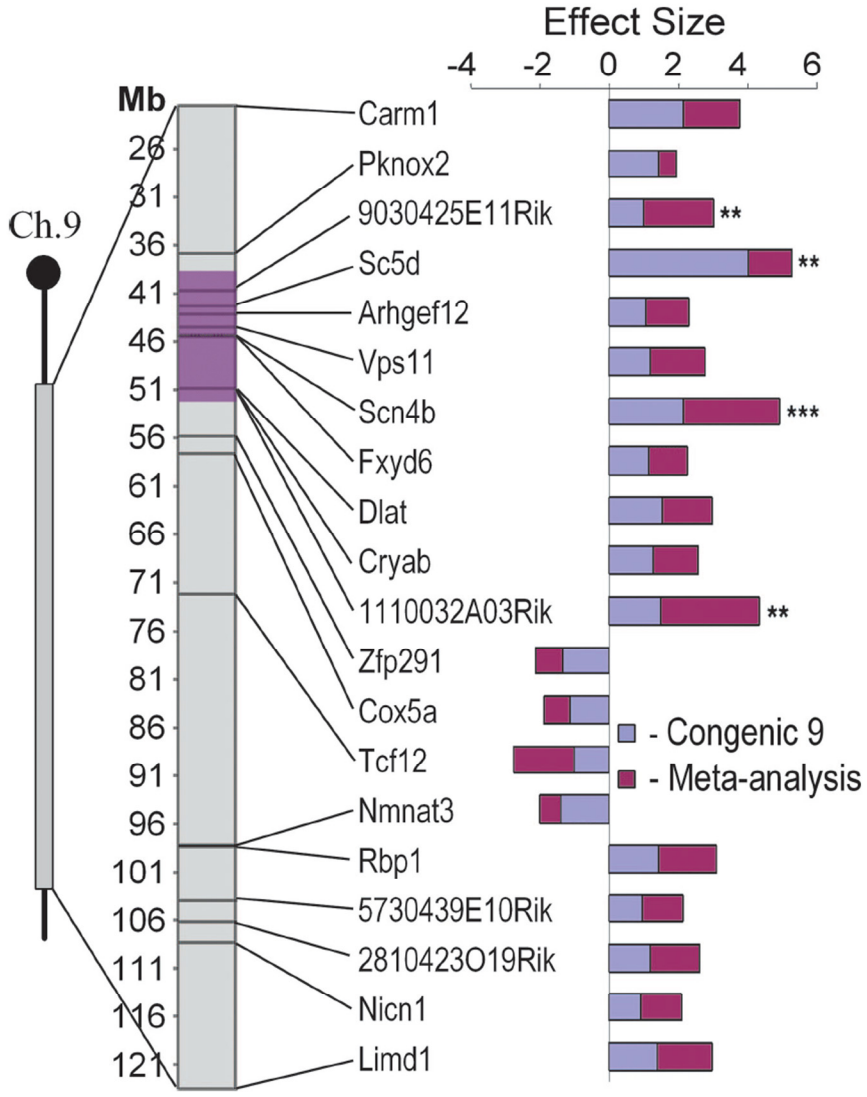


Figure 2.4: Candidate genes for alcohol preference QTL on mouse chromosome 9. Meta-analysis results filtered based on significant cis-regulation on chromosome 9. Asterisks indicate a correlation between gene expression and alcohol preference in a panel of BxD recombinant inbred strains (Phillips, Crabbe et al. 1994). Data from the WebQTL (www.genenetwork.org) Integrative Neuroscience Initiative an Alcoholism Brain mRNA M430 (April 2005 release) PDNN (Positional Dependent Nearest Neighbor) database (Wang, Williams et al. 2003). Originally published in Mulligan, Ponomarev et al. 2006.

CONCLUSIONS

Predisposition to excessive alcohol consumption is likely a key aspect of human alcoholism, but molecular determinants of this trait are difficult to study in humans. This study provides a microarray based meta-analysis of a behavioral phenotype using an effect size measure. The Cohen's *d* metric eliminated the need for standardization across array platforms and experiments allowing identification of consistently different transcripts between alcohol preferring and non-preferring genotypes, even if the transcriptional differences were small with respect to magnitude and/or significance.

The meta-analysis provided 5,182 targets, including 3,800 non-overlapping genes for further study. The significant differences between the high and low drinking genotypes represent numerous functional groups and a large range of cellular pathways as well as many genes (~25%) whose functions are not yet characterized. Thus, some genes whose functions are currently unknown are likely to be important determinants of alcohol consumption. The seventy-five genes with the largest effect size ($Q<0.05$, $|d| > 1.94$) (see figure 2.2A) fall into the broad categories of cellular homeostasis and neuronal function. Differences in the ability to maintain or reset homeostasis and adjust neuronal function in the brain likely underlie many aspects of alcohol responses and understanding these differences will be crucial for a better understanding of alcoholism.

The meta-analytical approach described in these results is designed to detect common differences found across the animal models used but will not detect differences specific for one model or one sex, nor will it detect mutations that affect alcohol preference for which no concomitant change in gene expression results. Unique but potentially important determinants of drinking require detailed analysis of each

dataset such as that which has been accomplished for the HAP/LAP selected lines (Tabakoff, Bhave et al. 2005) and for the B6.FVB hybrid that is described in chapter 4.

The meta-analytic approach utilizing an effect-size metric should be applicable to new datasets. Two-bottle choice alcohol drinking is the most widely used model of voluntary alcohol consumption but new models are emerging (Roberts, Heyser et al. 2000; O'Dell, Roberts et al. 2004; Finn, Belknap et al. 2005; Rhodes, Best et al. 2005). Follow up studies for candidate genes identified in previous array experiments have led to the understanding of molecular mechanisms for other behaviors (McClung and Nestler 2003; Yao, Gainetdinov et al. 2004) and is likely for the candidate genes detected by the meta-analysis as well.

In summary, this meta-analysis shows that distinct mouse models with genetic predisposition for elevated alcohol consumption have very consistent and reproducible differences in brain gene expression even though they have never consumed alcohol. Furthermore, the differences extend to thousands of transcripts forming many functional groups. Combining the meta-analysis with congenic approaches provides specific candidate genes for one alcohol preference QTL. These genes, as well as most of the key functional groups revealed by the current study were not anticipated by previous work, indicating both our ignorance of the molecular mechanisms driving alcohol consumption and, most importantly, the opportunities to reveal new mechanisms through large scale genomic screening approaches.

MATERIALS AND METHODS

Animal husbandry

All animals were adult (60-100 days old) mice, naïve to alcohol, with 24 hr *ad libitum* access to rodent chow, water, and 12 h :12 h lighting. All animals were housed and treated according to the NIH guidelines for the use and care of laboratory animals (Council 1996) and approved Institutional Animal Care and Use Committee protocols at each respective institution. Whole brain total RNA from naïve mice was used for all array analyses. Microarray hybridizations for the HAP/LAP lines, inbred strains and congenic 9 strains were completed on individual mice. The lines for high and low levels of short-term selection were hybridized with a pool of three samples for each represented N value (n = 10 mice per line). Details for each strain are listed below.

Short-term selection lines (STS)

Reciprocally crossed B6xD2 and D2xB6 F2 mice were between 8-10 weeks of age, housed individually and offered water in two 25-ml graduated cylinders fitted with stainless steel sipper tubes. Testing for 24-hour access two-bottle choice ethanol consumption followed the same methods used in other populations derived from the cross of C57BL/6J and DBA/2J inbred strains (Phillips, Crabbe et al. 1994; Phillips, Belknap et al. 1998). In the present study, B6D2 F1 mice were obtained from the Jackson Laboratories (Bar Harbor, ME) and bred to generate a founder population of B6D2 F2 mice (N=199, approximately equal males and females). The 10 male and 10 female mice with the highest consumption of 10% ethanol were bred to form the STDRHI2 line, while the 10 male and 10 female mice with the lowest consumption of 10% ethanol were bred to form the STDRLO2 line. In subsequent generations, the highest consuming mice of the STDRHI2 line and the lowest consuming mice of the STDRLO2 line were chosen to

perpetuate the lines. A total of 124-150 mice per generation were tested. Bidirectional selection continued similarly for each of three generations; the highest of the high line and the lowest of the low line were used to breed the next generation of high and low selection lines, respectively. Naïve mice with high and low levels of drinking from selection generation (S3) were used in the microarray experiment. Mice were 46-73 days old at the time of individual housing.

Mice were offered continuous access to two fluid-filled 25-ml graduated cylinders fitted with stainless steel sipper tubes. Initially, both were filled with tap water and the mice were given a 2-day acclimation period to get used to drinking from these tubes (days 1-2). One water tube was then replaced with a tube containing a 3% solution of ethanol in tap water (v/v). They had access to water vs. 3% ethanol for a 4 consecutive day period (days 3-6). The 3% ethanol tube was then replaced with a tube containing 10% ethanol in tap water (v/v). Consumption from the water and 10% ethanol tubes were measured for 4 consecutive days (days 7-10). Bottle positions were alternated every 2 days to disrupt any side-dependent preference in drinking. Food was distributed near both tubes to avoid food-related tube preference. Leakage/evaporation were measured from tubes with the same solution in them placed on cages not containing mice. The average volume lost from these tubes on 2-4 empty cages was subtracted from the volumes measured by meniscus reading. The selection index for drinking was based on the 10% concentration expressed in g/kg/day for each of the last days immediately prior to a bottle position change (i.e., the average of days 8 and 10).

High Alcohol Preferring/Low Alcohol Preferring lines (HAP/LAP)

Two replicate lines of HAP and LAP mice were originally generated from HS/Ibg mice (McClearn GE 1970) based on bidirectional selection outcome of two-bottle choice 10% alcohol solution vs. water consumption (Grahame, Li et al. 1999; Behm A 2003).

Male HAP1/LAP1 and HAP2/LAP2 mice, 70-90 days of age, were used from generations 26 and 19, respectively, for Affymetrix array analysis. Details of the gene expression studies which provide the data for the present analysis have been published elsewhere (Tabakoff, Bhave et al. 2005).

Inbred strains

Male BALB/cJ, C57BL/6J (B6), DBA/2J (D2) and LP/NJ inbred mice were raised and housed at the Veterans Affairs Medical Center in Portland, OR. LP/NJ mice were tested for two-bottle choice alcohol drinking in the laboratory of D. A. Finn and found to drink significantly less alcohol than C57BL/6J mice (unpublished data). Female B6, FVB/NJ (FVB) and FVBxB6 F1 mice were raised and housed at the University of Texas. F1 mice were recently found to drink significantly more alcohol than the progenitor strains when tested in an accelerating concentration two-bottle choice paradigm (Blednov, Metten et al. 2005).

Congenic strain

A D2 QTL region for alcohol preference from 9 to 58 cM on chromosome 9 was introgressed onto a B6 genomic background through numerous generations ($\geq$ 12) of breeder selection by Massachusetts Institute of Technology microsatellite marker genotyping and backcrossing to B6. The B6.D2^{Ch9} congenic chromosome 9 mice were previously tested for alcohol two-bottle choice drinking and found to drink significantly less than B6 control mice (difference $\approx$25%, T.J.Phillips, unpublished data). Whole brain mRNA was isolated from jointly housed and age-matched (70-100 days) congenic and B6 mice.

Oligonucleotide Microarrays

Whole brain samples containing at least 10 μg of total RNA were hybridized precisely following manufacturer's protocol (Affymetrix, Santa Clara, CA). Methodology for arrays completed for HAP/LAP mice have been previously published (Tabakoff, Bhave et al. 2005). RNA from STS mice and mice from the inbred strains BALB/cJ, C57BL/6J, DBA/2J, and LP/NJ were hybridized to Affymetrix 430 A and B chips and analyzed at the Oregon Health Sciences University Gene Microarray Shared Resource Facility. Total RNA was isolated with TRIZOL7 Reagent (Life Technologies Inc., Gettysburg, MD) using a modification of the single-step acid guanidinium isothiocyanate phenol-chloroform extraction method, according to the manufacturer's protocol. The extracted RNA was then purified using RNAeasy (Qiagen Inc, Valencia, CA) and quantified by UV spectroscopy. All samples were hybridized to the GeneChipTest3 for quality control. Only samples that had a 260/280 ratio >1.8, good visual integrity on ethidium bromide stained denaturing formaldehyde agarose gels and that passed target performance recommended thresholds, were used. For further details, see: http://www.ohsu.edu/gmsr/amc. Robust microarray average (RMA) and position dependent nearest neighbor (PDNN) were both used to analyze the array data. Results are only presented for PDNN; however, there was no qualitative difference between the RMA and PDNN data. RMA was implemented using the Affymetrix package (version 11/24/2003) from the Bioconductor version 1.3 framework (http://www.bioconductor.org) running within the R (version 1.8.1) programming environment for both Windows and Solaris (http://cran.stat.ucla.edu). PDNN analysis (Zhang, Miles et al. 2003) using default settings was implemented using PerfectMatch software (v.2.3) (http://idub.mdacc.tnc.edu/~zhangli/PerfectMatch). Both programs are available for download without charge.

Custom cDNA arrays

cDNA fragments PCRed from clones were printed on poly-L-lysine (Sigma, St. Louis, MO) coated microscope slides (Erie Scientific, Portsmouth, NH) using a custom built robotic arrayer as previously described (DeRisi, Iyer et al. 1997). The clones were from several cDNA libraries, including expressed sequence tags cloned in the Bergeson laboratory; Research Genetics/Invitrogen (Carlsbad, CA) clone sets, BMAP (brain molecular anatomy project) and SV (sequence verified); and NIA (National Institute on Aging, (Tanaka, Jaradat et al. 2000)) clone sets 7.4K and 15K. Microarray hybridizations were performed using the Array 350 microarray labeling kit according to manufacturer's protocol (Genisphere, Inc., Hatfield, PA). Extracted total RNA samples from whole brain were reverse transcribed, labeled with the Cyanine-3 (Cy-3) fluorophore, and hybridized against a common reference RNA labeled with Cyanine-5 (Cy-5). The common reference is from whole brain RNA extracted from 100 male C57BL/6J (Jackson Laboratories, Bar Harbor, ME) mice that were 70 days old. All arrays contained the same reference RNA in the Cy-5 channel and were normalized using within-print tips lowess non-linear normalization (33). The microarrays were scanned using the GenePix 4000B Microarray Scanner, gridded using the software package GenePix v 5.1 (Axon Instruments, Sunnyvale CA) and normalized using within-print tips lowess non-linear normalization (Workman, Jensen et al. 2002). Normalized array data were stored using the Longhorn Array Database (LAD) (Killion, Sherlock et al. 2003) and then standardized using the red channel (common reference RNA) as the baseline standard using software developed in the Bergeson lab (Perl programs created by A.E.Berman).

Meta-analysis

Five data sets listed as 1-5 in Table 1 were used for meta-analysis. Alcohol-preferring and non-preferring mice were first compared using a parametric statistical test.

Student's *t*-test for independent samples was used to compare "high" and "low" drinking selected lines (studies 1-3). Both sets of inbred strains (studies 4 and 5) were analyzed independently by an *F* test using single contrasts. Contrast analysis is a type of analysis of variance, which, compared to omnibus ANOVA, is best suited to test specific hypotheses when 3 or more groups are to be compared. For each study that used inbred strains, alcohol-preferring genotypes were compared (contrasted) to alcohol non-preferring genotypes using weighting coefficients. For example, for study 4, the C57BL/6J (B6) strain was assigned a weighting coefficient of 3 and then contrasted to a combination of three other strains, each assigned a weighting coefficient of (-1). The resulting *t* and *F* values for each transcript were then used to estimate effect size by calculating a Cohen's *d* statistic, which is the difference between the groups expressed in pooled within-group standard deviation units. The following formulas were applied: $d = 2t/\sqrt{df} = 2\sqrt{F/df}$, where *df* represents the degrees of freedom. The direction of change was coded in the resulting Cohen's *d* values so, that <positive> values indicated an up-regulation and <negative> values, down-regulation of transcripts in alcohol preferring compared to non-preferring genotypes.

One potential problem of meta-analytic approaches is the generalization of differences among the individual studies. Our analysis includes data from selectively bred lines and inbred strains. Although both genetic approaches have been successfully used to study complex traits, no systematic efforts have been made to compare the two methods with respect to their power to detect genetic correlations. However, there is a general agreement that comparison of two selected lines is much more powerful than any single comparison of two inbred strains. Because the selected line approach uses information from individual animals while the inbred strain method uses strain means, more inbred strains are generally needed to study genetic correlations (Crabbe, Phillips et

al. 1990). Therefore, in an attempt to equate the genetic contribution to ethanol drinking across studies and to minimize over-generalization that could come from the small number of inbred strains in each individual data set, data from the two inbred strain studies (4 and 5 in Table 1) were collapsed by averaging Cohen's *d* values for each transcript. To do so, Affymetrix and cDNA platforms were matched by gene symbols. If a gene was represented by more than one transcript on the cDNA array, the transcript with greater absolute effect size was used for the combined data set. To minimize an outlier effect, extremely deviant effect sizes were adjusted to a maximum absolute value of $d = 4$. Finally, the four Cohen's *d* values generated for the three pairs of selected lines and the combined set of inbred strains were averaged and a *z*-test was used to test significance of deviation of the mean effect size from zero. QVALUE software (http://faculty.washington.edu/~jstorey/qvalue) was used to estimate false discovery rate (*Q*) for the meta-analysis results (Storey and Tibshirani 2003), which estimates the proportion of all declared significant results that are expected to be false positives.

Congenic 9 data filter

We used microarray data obtained from the B6.D2^{Ch9} congenic strain to filter transcripts detected by meta-analysis as regulated significantly between ethanol preferring and non-preferring genotypes and located within the introgressed region known to contain an ethanol preference QTL on mouse chromosome 9. Specifically, chromosome 9 transcripts significantly regulated between the congenic and control lines ($p<0.05$ by a *t*-test) were identified and matched for direction of change with data from meta-analysis. Only genes detected by both approaches were selected as putative candidates for *cis*-regulation of alcohol preference.

Transcription Factor Binding Site Overrepresentation Analysis

The oPOSSUM database (Ho Sui, Mortimer et al. 2005) was used to analyze transcription factor binding sites (TFBS) in genes that were either significantly (Q<0.05 and $|d| \geq 0.05$) up- or down-regulated in alcohol-preferring mice compared to non-preferring mice. The two lists had 2,388 genes with 1,011 recognized as orthologs in oPOSSUM and 1,580 genes with 546 recognized for up- vs. down-regulation respectively. Overrepresented TFBS were determined for each group of genes based on one-tailed Fisher exact probabilities (p-value) and z score ranking. Determination of a non-random association of the TFBS uses the p-value for a comparison of the proportion of target genes (co-expressed genes) containing a specific TFBS to the proportion of background genes (all genes represented by the microarray) for that TFBS. The number of times a TFBS appears in a promoter region (2000 bases upstream of the transcription start site) is irrelevant. The Z-score, however, compares the rate of every occurrence of a TFBS in target genes with the rate of each occurrence in background genes. Following the initial analysis, the genes for the transcription factors themselves were individually checked to see whether: 1) they were present on the arrays, and 2) transcription differences were consistent with the promoter TFBS analysis.

Functional Overrepresentation Analysis

Transcripts that pass the statistical thresholds (Q<0.05 and $|d| \geq 0.05$) and had an assigned gene symbol were annotated into functional groups using the WebGestalt (web-based gene set analysis toolkit http://genereg.ornl.gov.webgestalt). Functional annotations based on Gene Ontology (GO) Consortium (www.geneontology.org), BioCarta pathways (http://www.biocarta.com/genes/index.asp), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/keg) were carried out. The WebGestalt database was then used for functional over-representation analysis. This

database uses a hypergeometric test to compare representation (number of significantly regulated genes) of a certain functional group to representation of this functional group on a background list (all genes from the meta-analysis that passed the array filtering criteria). In other words, it estimates whether a certain biological group or pathway is active (*i.e.* represented by more significant genes than expected by chance). All 3,800 genes detected by meta-analysis were used to determine over-represented pathways in KEGG and BioCarta. Only the 500 genes with the largest effect sizes from this list were used for similar GO-based analysis due to processing limitations in the WebGestalt database.

Chapter 3: Alcohol Trait and Transcriptional Genomic Analysis of C57BL/6 Substrains

INTRODUCTION

Inbred mouse strains represent a valuable genetic model system for the investigation of complex behaviors. The availability of many different inbred strains that differ in a range of phenotypes coupled with the development of molecular genetics techniques that allow for targeted gene manipulation and transcriptome analysis has expedited the search for genes that control specific traits. Among the most commonly studied and most widely utilized inbred strain is C57BL/6 (B6). Like many of the modern inbred mouse strains, B6 was originally derived from stocks kept in the early 1900's by geneticists and mouse fanciers alike (Beck, Lloyd et al. 2000). C57BL/6 was initially isolated from the C57BL/10 strain by 1937 and the C57BL/6J (B6J) colony was founded at Jackson Laboratories in 1948. At least 40 subsequent sub-colonies have since been derived from B6J between 1930 and 1970 (see: JAX Notes™ 498, http://jaxmice.jax.org/library/notes/498a.html) including C57BL/6N (B6N), a sub-strain founded by the National Institute of Health (NIH) from B6J in 1951. In 1974 Charles River Laboratories started their own sub-colony C57BL/6NCrl (B6C) from the 32nd generation of the B6N colony. Thus, the B6J and B6C substrains have arisen through the propagation of new B6 colonies by different breeding facilities, including at The Jackson Laboratory and Charles Rivers Laboratory, and have been inbred separately for about 5 decades. Although derived from the same original B6 inbred mouse strain, separation of B6 colonies leads to genetic divergence over time. Within isolated sub-colonies, genetic drift can be noticeable as soon as 20 generations after separation (Bailey, 1977, Bailey, 1982). If a sub-colony is founded from an inbred strain prior to the 40th generation of inbreeding, genetic drift may occur due to remaining heterozygosity within the original

inbred strain. Indeed, SNP analysis of B6 substrains supports the existence of residual heterozygosity despite extensive inbreeding at the time the substrains were separated and also raises the possibility that some loci may be under selective pressure to retain allelic heterozygosity even after many generations of inbreeding (Petkov, Ding et al. 2004). Of course, genetic drift can also occur if spontaneous gametal mutations become fixed within the subcolony and several examples of such mutational events have recently been identified between B6J and B6C. For instance, a spontaneous deletion occurring only in the B6J strain resulted in the loss of several exons of the nicotinamide nucleotide transhydrogenase (*Nnt*) gene (Huang, Naeemuddin et al. 2006). Additionally, from a panel of 867 microsatellite markers, 13 were found to be polymorphic between B6J and B6C (Hovland, Cantor et al. 2000). Phenotypically, the two B6 substrains differ in a wide range of traits including cardiac function during anesthesia and sensitivity to anesthetics (Roth, Swaney et al. 2002), body weight (Green, Singh et al. 2007) aspects of the fear response (Radulovic, Kammermeier et al. 1998; Siegmund, Langnaese et al. 2005), and some alcohol-related traits (Green, Singh et al. 2007, Khisti, Wolstenholme et al. 2006, Ramachandra, Phuc et al. 2007).

In the alcohol research field, B6J mice supplied by The Jackson Laboratory have been one of the most abundantly used laboratory models, most likely due to the fortuitous discovery early in the history of the substrain that B6J mice have a high alcohol preference phenotype (McClearn and Rodgers 1959). B6J mice have since been shown to drink more alcohol than other commercially available inbred mouse strains (Belknap, Crabbe et al. 1993). The phenotype of high alcohol preference in the B6J strain has remained quite stable since it was first reported, likely due to the fact that preference for alcohol is a complex trait controlled by many genes of varying but smaller effect size (Wahlsten, Bachmanov et al. 2006). Unfortunately, most of the genes and molecular

mechanisms involved in driving initial alcohol preference have yet to be elucidated. In 1982 researchers discovered a large difference in alcohol preference between B6J and a now extinct B6 substrain derived from B6N and maintained by Simonsen laboratories (Blum, Briggs et al. 1982). More recently, it was discovered that B6J also differed from B6C for several alcohol related phenotypes, including sensitivity to fetal alcohol syndrome (Green, Singh et al. 2007), alcohol consumption after ethanol deprivation (Khisti, Wolstenholme et al. 2006) and ethanol preference and dopamine release in the ventral striatum after exposure to alcohol (Ramachandra, Phuc et al. 2007).

It is generally accepted that B6J and B6C have a very similar genetic background with the apparent exception of several discrete loci, which results in substrain phenotypic divergence for several traits, including alcohol preference. In order to provide researchers with crucial information that may prove useful to gain insight into the molecular mechanics of phenotypic differences observed between these two substrains now and in the future, we completed brain region transcriptome analyses on alcohol naïve B6J and B6C.

RESULTS

Alcohol consumption and alcohol preference are lower in the B6C substrain

Consumption and preference for alcohol were significantly lower in both female and male B6C mice regardless of whether the mice were purchased and shipped from their respective vendors or bred in house. In an attempt to recapitulate the Simonson findings of their extinct B6 substrain in our laboratory using the B6J and B6C substrains, vendor purchased mice were tested for two-bottle choice alcohol drinking using a scaled increase in % alcohol starting at 3%. Both strains showed a strong preference for alcohol but as shown in figure 3.1A, intake of alcohol was found to be significantly higher in female B6J compared to B6C mice (p=0.001, between groups repeated measure ANOVA). Consumption of alcohol by both substrains increased linearly with concentration until leveling off at 12% alcohol and above, at which point B6C consumed approximately 25% less than B6J. Average consumption of alcohol for B6C and B6J was 16.0 g/kg/day (SEM±1.91) and 21.8 g/kg/day (SEM±1.02) respectively. The B6 mice showed the largest substrain difference at the 15% alcohol solution for which B6J consumed >40% more alcohol than B6C. Of note is that B6C showed a reduced range of preference over increasing alcohol solutions compared to B6J (figure 3.1B) and B6C had a significantly lower preference for alcohol compared to B6J (p<0.001, two-way ANOVA). The B6C mice only showed a preference for the 6% and 12% alcohol solutions while B6J showed preference for the 3% through 15% alcohol solutions. In fact, B6C actually demonstrated mild alcohol avoidance for the 3%, 18% and 21% solutions of alcohol, in contrast to B6J, which never demonstrated avoidance at any of the concentrations measured. To ensure that the differences observed between the substrains were not simply environmental influences, mice from each substrain were bred in house. As observed for mice purchased from different vendors, both male and female

B6JUT mice consumed significantly more alcohol than their male and female B6CUT counterparts ($p<0.05$, *t*-test) (figure 3.1C) and both male and female B6JUT mice demonstrated a significantly higher preference for alcohol ($p<0.05$, *t*-test) than B6CUT mice of both sexes (data not shown). Consistent with other studies of alcohol consumption in mice, female mice consumed more alcohol than male mice.

In order to determine if the difference observed in alcohol consumption and preference was due to discrepancies in taste between the two B6 substrains, animals purchased from the vendor were evaluated for preference of sweet and bitter solutions because alcohol is thought to represent a combination of both tastes. No difference in percent preference was observed between B6J and B6C mice for 0.033% saccharine (sweet) and 0.03 mM quinine (bitter) solutions (87.0±7.0, 87.5±4.2, p=0.6; 36.8±4.5, 44.7±4.06, p=0.2 respectively. Mean ± SEM, n=10).

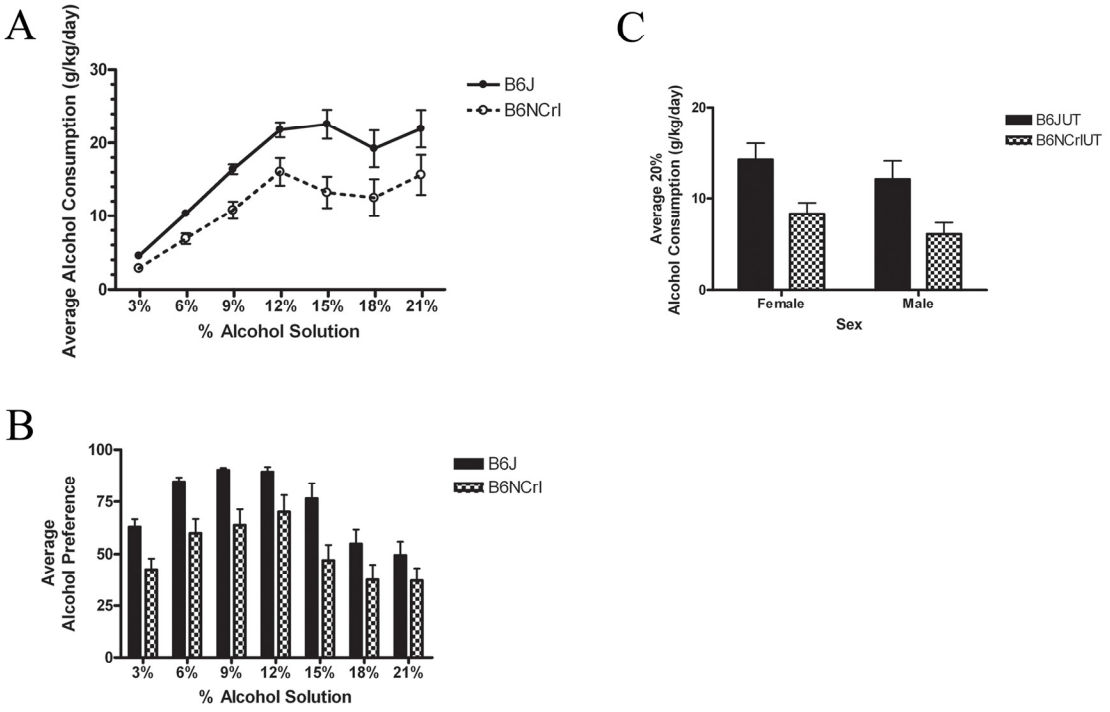


Figure 3.1: Alcohol consumption and preference levels between the B6 substrains. (A) Average alcohol consumption of each strain is plotted against alcohol concentration. The dashed line represents female mice purchased from Charles River Laboratories (B6NCrl) and the solid line represents female mice purchased from Jackson laboratories (B6J). The effect of strain is significant (p=0.001, ANOVA). (B) Mean alcohol preference between substrains is shown for each solution of alcohol. Female mice from Jackson Laboratory (black) have a significantly higher preference for alcohol compared to Charles River mice (checkered) (p<0.001, ANOVA). (C) Male and female B6J mice bred in house (black bar) consume significantly more alcohol than their same sex Charles River B6NCrl counterparts bred in house (checkered) counterparts (p<0.05).

No strain difference detected for alcohol-induced loss of righting reflex

Initial sensitivity to alcohol may affect subsequent drinking. Therefore, alcohol-induced loss of righting reflex was measured to gain insight as to whether a negative response or differences in alcohol metabolism were contributing factors mediating the substrain difference in alcohol consumption and preference. No difference was observed between vendor purchased B6J or B6C for the latency to lose the righting reflex, the duration of the loss of the righting reflex or in the BEC at the time of recovery of the righting reflex (figure 3.2). Taken together, these findings imply no difference in initial sensitivity or acute tolerance to the hypnotic effects of ethanol and no difference in alcohol metabolism between the substrains. In agreement with this data, a recent study found no significant difference in ethanol clearance between substrains (Ramachandra, Phuc et al. 2007).

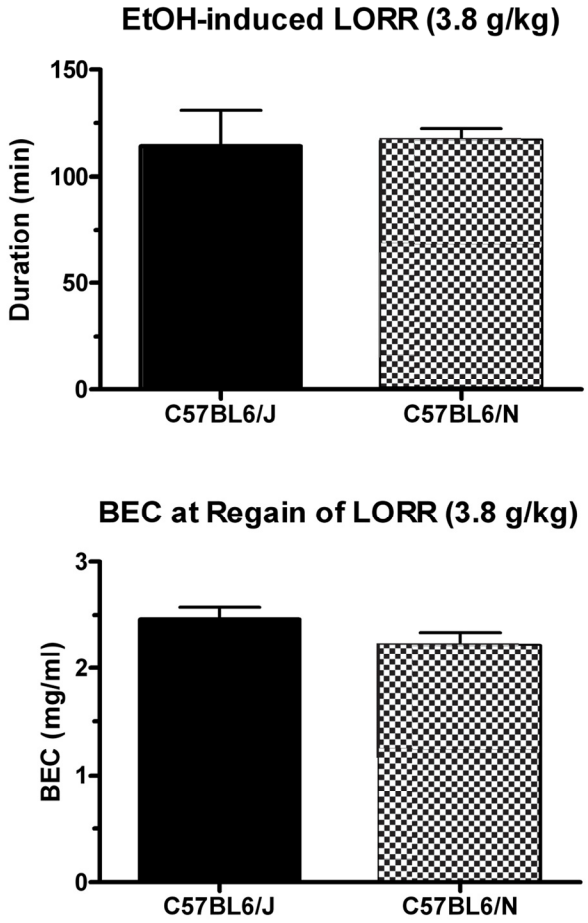


Figure 3.2: Initial sensitivity to the hypnotic effects of alcohol in vendor purchased B6 substrains. There is no difference in initial sensitivity as measured by the loss of righting reflex (LORR). No difference was observed between substrains for the duration of the loss of the righting reflex (116.8±5.4 min and 114±16.9 min, respectively, p=0.9, n=10) or in the BEC at the time of recovery of the righting reflex (2.2±0.1 mg/ml and 2.5±0.1 mg/ml, respectively, p=0.1, n=10).

Analysis of correlated transcription across brain regions revealed significant divergence at the genomic level

A meta-analytic approach was used to identify transcripts that were coordinately changed by a large degree between substrains across all 4 brain regions. This approach revealed 86 unique transcripts significantly co-regulated (z-test $p < 0.01$, average $|d| \geq 0.8$ and average fold change $\geq |1.5|$) between B6J and B6NCrl across all four brain regions (figure 3.3, table 3.1 and table 3.2). Although this study was not designed to identify and differentiate primary substrain differences from those secondary to upstream genetic polymorphisms, genes that share large changes in expression across brain regions likely result from genetic changes between the substrains and several bioinformatics approaches were used to gain insight into the brain transcriptome differences between the B6J and B6C mice.

An analysis of chromosomal location of the co-regulated genes revealed that two of the genes, *D14Ertd449e* and *Plac9*, were located adjacent to each other on chromosome 14. Visualization of this region using the Ensembl Genome Browser (www.ensembl.org) and NCBI MapViewer (www.ncbi.nlm.nih.gov/mapview) revealed potential duplication events of several genes within the B6J genome including an apparent 3X copy number in the B6J strain which is the reference strain for both websites (figure 3.4A). To gain insight as to whether the observed change in gene copy number was a B6 specific phenomenon rather than normal across other inbred strains of mice (i.e. an orthologous event), we examined the expression of representative genes from the chromosome 14 region within our own inbred strain microarray datasets as well as within publicly available inbred strain microarray datasets (www.genenetwork.org). *D14Ertd449e* and *Plac9* were the only genes from within the chromosome 14 region that were represented on our microarray and we found strain differences for these genes

consistent with potential duplication events in the B6J genome or, in other words, a paralogous event since, at least in humans, there is no aberrant change in copy number (figure 3.4B). A large decrease in average whole brain expression for both the *Plac9* and *D14Ertd449e* loci were found for, FVB/NJ (FVB) and SJL/J (SJL) and, to a lesser extent, for DBA/2J (D2) when compared to B6J. In agreement with the expression data, two separate comparative genomic studies had identified the same chromosome 14 region as containing copy number gains in B6J and DBA/2J (D2) that were not present in FVB/NJ (FVB) or SJL/J (SJL) (Snijders, Nowak et al. 2005; Graubert, Cahan et al. 2007). The change in expression of *Plac9* and *D14Ertd449e* between substrains was confirmed by quantitative real time PCR and confirmation of the genomic change in copy number is in progress (figure 3.3C). Analysis of the chromosomal regions for the other top divergent genes revealed that the chemokine ligand 21b (*Ccl21b*) and the Riken gene, *4933409K07Rik*, which both occupy a small region on chromosome 4 also appeared to have a higher copy number in the B6J genome (an apparent 2X copy number) compared to B6NCrl and several inbred strains of mice (figure 3.5).

Since expression can change due to events downstream of a genetic polymorphism, we looked for potential *cis*-regulation or enhancer activity using insight from a combination of chromosomal localization information and coordinated transcription factor binding site (TFBS) data. We developed a novel technique to detect chromosomal regions that contained clusters of co-localized and co-expressed genes that showed significant expression differences between substrains across brain regions using a binomial distribution comparison method. Using this method, regional chromosome overrepresentation was detected on four chromosomes and the results are shown in figure 3.6. Although some of the genes in each of the 4 regions have a few TFBS in common,

none of the identified chromosomal regions has a common TFBS shared by all the genes within a cluster (data not shown).

Next, we looked for pathway coordination in the gene expression differences using the GOTree function in WebGestalt. Interestingly, nearly 40% of the genes were of unknown function but overrepresented pathways and cellular localizations were detected including: signal transduction (*Rapgef, Eng, Rasa2*; p=0.04), phosphorylation (*Rapgef, Ilk, Mapk4k4, Rlf, Prkd2*; p=0.02), transition metal ion binding (*Mmp14, Galnt2, Hic2, Usp27x, Hmox2, Tnfaip3, Trim10, Rasa2, Rlf, Prkd2*; p=0.03), cellular lipid metabolism (*A4galt, Ptges, B3gnt5, Smpd3*; p=0.03), peptidase activity (*Psen1, Mmp14, Usp27x, Tnfaip3, Rlf*; p=0.02), UDP-glycosyltransferase (*A4galt, Galnt2, B3gnt5; p*=0.001), localization to golgi (*Psen1, A4galt, Galnt2, Smpd3; p*=0.006), and nuclear envelope (*Psen1, Ptges, Tpr; p*=0.002). In addition, two small, related categories were notable for their potential to contribute to brain survival: positive regulation of cell proliferation (*Cd28, Ilk; p*=0.04) and anti-apoptosis (*Psen1, Cd28; p*=0.02).

CLID	GBID	Entrez	Name	Symbol	Chr	CoStr p	CoStr FC	Hip p	Hip FC	Vb p	Vb FC	Cer p	Cer FC	AVG FC	CoStr d	Hip d	Vb d	Cer d	AVG d	p_Zhst	J CoStr AVG	J Hip AVG	J Vb AVG	NCrl Vb Avg	J Cer AVG	NCrl Cer AVG
H3116P05	BG071178	17118	Mdm 1	Mdm1	X	1.63E-07	-3.66	1.24E-05	-3.41	3.70E-06	-3.05	-3.63	-3.11	-3.46	5.11	3.08	3.06	3.93	3.56	1.83E-08	5.39	3.15	2.87	5.10	4.67	5.23
6212B1	AI451073	320172	RIKEN cDNA E230016M11 gene	E230016M11Rik	X	4.09E-02	-2.18	1.69E-02	-3.55	6.12E-01	-1.41	-1.41	-2.69	-2.06	0.38	0.81	1.14	1.06	0.28	1.14E-04	5.29	3.05	2.87	5.10	4.67	5.23
H4018D09	BQ552820	74340	RIKEN cDNA 4631427C17 gene	4631427C17Rik	6	1.98E-02	-2.06	7.72E-03	-4.96	3.93E-02	-3	3.77E-01	-1.2	-2.81	1.41	1.66	1.21	0.49	1.19	2.24E-06	5.63	4.58	6.61	6.35	4.76	4.43
A1838322	AI838322	18829	Chemokine (C-C motif) ligand 21b	Ccl21b	4	1.91E-03	-4.31	6.61E-01	-1.26	1.01E-01	-2.21	1.50E-03	-3.22	-2.75	2.04	0.24	0.94	2.10	1.33	3.19E-03	7.62	5.52	5.72	5.39	5.76	4.62	6.87	5.18
H3017C07	BG064202	270701	Tripeptidase like 1	Tpl1	4	2.51E-04	-3.88	1.91E-05	-2.44	2.87E-07	-2.65	6.91E-09	-3.47	-2.63	2.59	2.94	4.69	6.57	4.06	1.93E-08	8.27	7.59	8.69	9.35	7.93	8.82	7.02
A1839337	AI839337	211623	Placenta specific 1	Plac9	14	2.20E-08	-2.63	1.72E-08	-2.76	8.93E-08	-2.3	1.34E-09	-2.68	-2.60	6.00	6.33	5.36	7.45	6.28	0.00E+000	8.55	7.16	8.67	7.19	8.86	7.66	8.24	6.62
H4049A06	BQ557910	12831	Procollagen, type V, alpha	Col5a1	2	7.52E-02	-2.38	3.77E-02	-3.98	2.29E-01	-1.83	3.08E-02	-1.6	-2.45	1.03	1.23	0.67	1.28	1.05	2.82E-14	5.21	3.96	5.12	4.12	5.70	4.83	4.73	4.05
H4049C12						1.22E-02	-2.38	3.92E-02	-1.63	8.99E-01	-1.18	3.12E-02	-3.12	-2.35	0.88	1.06	0.29	1.28	0.93	6.23E-05	5.14	3.92	4.95	3.69	4.06	3.92	5.58	3.65
H3006C03	CK334720	72415	Shugoshin-like 1 (S. pombe	Sgol1	17	6.81E-03	-4.48	1.09E-01	-1.65	7.16E-01	-1.23	8.96E-02	-2.03	-2.35	1.69	0.91	0.20	0.97	0.95	1.97E-03	4.86	2.70	4.83	4.11	4.61	4.32	4.92	3.90
H3036G05	BG066172	170942	Erythroid differentiation regulator	Edr1	3	4.80E-03	-2.34	1.66E-03	-2.3	4.54E-03	-1.79	1.99E-03	-2.95	-2.35	1.79	2.07	1.60	2.03	1.92	0.00E+000	10.83	11.79	10.59	11.78	10.94	12.74	11.19
5982B9	AI447333		Transcribed locus		3	7.59E-02	-2.19	1.39E-02	-3.48	8.91E-01	1.06	8.09E-02	-2.8	-2.35	1.02	1.50	-0.07	1.01	0.88	9.54E-03	3.69	2.55	4.70	2.25	2.69	2.78	3.86	2.37
H4010H01	BQ551677	83703	Debranching enzyme homolog 1 (S. cerevisiae)	Dbr1	9	7.99E-03	-2.53	4.00E-02	-2.54	6.42E-01	-1.27	2.11E-02	-3.01	-2.34	1.65	1.21	0.25	1.39	1.13	2.18E-04	4.52	3.18	5.06	3.72	3.27	2.92	5.14	3.55
A1853889	AI853889		Transcribed locus		15	7.04E-03	-2.01	6.34E-04	-2.11	4.74E-07	-2.78	4.34E-05	-2.45	-2.34	1.69	2.33	4.68	3.13	2.98	4.70E-06	7.70	6.69	7.74	6.67	8.42	6.95	7.65	6.36
H3078C03	BG073934	85675	Bicaudal C homolog 1 (Drosophila	Bicc1	10	1.21E-01	-1.59	3.71E-03	-1.99	2.89E-05	-3.28	8.09E-03	-2.12	-2.25	0.88	1.66	3.24	1.65	1.91	1.05E-04	7.03	6.36	7.34	6.36	7.64	5.92	7.96	6.87
A1854596	AI854596	16202	Integrin linked kinase	Ilk	7	5.59E-03	-1.48	2.22E-02	-3.53	1.00E-01	-2.02	2.19E-01	-1.91	-2.24	1.37	0.94	0.69	1.19	0.98	4.02E-07	7.38	5.62	6.64	6.00	4.98	5.74	4.81
A1853531	AI853531	74155	ERBB receptor feedback inhibitor	Errfi1	4	2.05E-01	-2.05	3.55E-01	-1.38	2.31E-01	-1.35	6.62E-03	-4.09	-2.22	0.71	0.51	0.67	1.69	0.90	8.47E-04	4.67	3.63	4.81	4.35	4.86	4.44	5.18	3.15
H4048B09	BQ557711	68814	RIKEN cDNA 1200014J11 gene	1200014J11Rik	11	2.56E-01	-2.45	2.53E-01	-2.42	2.53E-02	-2.36	1.48E-02	-3.53	-2.19	0.63	0.64	1.34	1.48	1.02	3.98E-06	5.68	4.39	4.66	3.39	4.59	3.35	4.46	3.86
H3146G09	BG075404	54651	Ubiquitin specific peptidase 27, X chromosome	Usp27x	X	1.22E-01	-3.03	1.54E-01	-1.69	2.29E-02	-2.53	8.70E-02	-1.44	-2.17	0.88	0.81	1.37	0.98	1.01	6.66E-16	6.45	4.85	5.52	4.76	5.68	4.34	4.88	4.36
H4059F03	BQ559660	108105	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1	B3gnt5	16	9.08E-02	-2.2	7.13E-01	-1.36	1.34E-02	-3.02	1.56E-02	-2.06	-2.16	0.97	0.20	1.51	1.47	1.04	6.67E-04	4.71	3.57	4.06	3.62	5.84	4.25	5.05	4.01
H3054H10	BG080549	66039	DNA segment, Chr 14, ERATO Doi 449, expressed	D14Ertd449e	14	9.82E-04	-1.88	1.59E-04	-2.27	9.54E-04	-2.07	7.75E-04	-2.36	-2.15	2.22	2.73	2.23	2.28	2.36	0.00E+000	7.72	6.81	7.92	6.74	8.12	7.07	7.70	6.46
H4078C12	BQ562850					6.99E-02	-3.09	3.81E-02	-3.56	2.27E-02	-2.67	9.95E-01	1	-2.08	1.05	1.22	1.37	0.00	0.91	3.48E-03	4.95	3.32	4.91	3.08	4.34	2.92	4.06	4.06
H3108G07	BG072279	66039	DNA segment, Chr 14, ERATO Doi 449, expressed	D14Ertd449e	14	1.85E-07	-2.23	1.11E-04	-1.78	8.52E-07	-1.76	7.88E-07	-2.28	-2.01	5.06	2.84	4.45	4.48	4.21	0.00E+000	9.81	8.65	9.59	8.76	9.83	9.02	9.63	8.44
H3103F09	BG071832	67968	Adducin 3 gamma	Add3	19	3.06E-01	-1.98	2.08E-01	-1.76	2.18E-01	-1.69	4.42E-03	-1.57	-1.81	0.57	0.71	0.69	1.81	0.94	1.18E-03	4.44	3.45	4.24	3.41	4.40	3.48	4.76	4.11
H4078E03	BQ562876	14017	Ecotropic viral integration site 2.	Evi2a	11	2.74E-02	-2.56	1.55E-01	-1.3	2.41E-02	-1.65	6.70E-03	-1.7	-1.80	1.32	0.80	1.35	1.70	1.29	2.52E-12	5.72	4.36	6.39	6.02	6.58	5.86	6.32	5.55
A1851309	AI851309	329867	Transmembrane 106c	Tmem106c	7	5.42E-02	-3.87	3.02E-01	-1.2	2.06E-02	-2.11	4.16E-01	-1.23	-1.88	1.12	1.42	0.64	0.45	0.91	4.57E-05	5.39	3.58	5.18	4.89	5.53	4.64	5.96	4.94
H3072D08	BG069071	67464	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	14	2.50E-04	-1.88	4.41E-06	-1.73	4.15E-06	-1.78	1.18E-07	-2.07	-1.87	2.60	3.86	3.88	5.25	3.90	5.46E-13	10.47	9.56	10.12	9.32	10.32	9.49	10.21	9.16
A1850133	AI850133	223864	Rap guanine nucleotide exchange factor (GEF) 3	Rapgef3	15	7.45E-01	-1.04	1.26E-02	-2.65	2.67E-01	-1.46	2.66E-02	-2.17	-1.83	0.18	1.53	0.62	1.32	0.91	3.61E-03	7.15	7.10	8.28	6.88	7.97	7.42	7.42	6.30
H4044F09	BQ557208					3.00E-01	-2.31	2.83E-01	-1.79	1.62E-02	-1.93	3.04E-01	-1.29	-1.83	0.57	0.60	1.46	0.57	0.80	2.76E-04	4.95	3.74	5.36	4.52	5.24	4.29	5.32	4.95
H4044G01	BQ557215	27360	Adducin 3 gamma	Add3	19	3.06E-01	-1.98	2.08E-01	-1.76	2.18E-01	-1.69	4.42E-03	-1.57	-1.81	0.57	0.71	0.69	1.81	0.94	1.18E-03	4.44	3.45	4.24	3.41	4.40	3.48	4.76	4.11
H4078E03	BQ562876	14017	Ecotropic viral integration site 2.	Evi2a	11	2.74E-02	-2.56	1.55E-01	-1.3	2.41E-02	-1.65	6.70E-03	-1.7	-1.80	1.32	0.80	1.35	1.70	1.29	2.52E-12	5.72	4.36	6.39	6.02	6.58	5.86	6.32	5.55
A1851309	AI851309	329867	Transmembrane 106c	Tmem106c	7	5.42E-02	-3.87	3.02E-01	-1.2	2.06E-02	-2.11	4.16E-01	-1.23	-1.88	1.12	1.42	0.64	0.45	0.91	4.57E-05	5.39	3.58	5.18	4.89	5.53	4.64	5.96	4.94
H3072D08	BG069071	67464	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	14	2.50E-04	-1.88	4.41E-06	-1.73	4.15E-06	-1.78	1.18E-07	-2.07	-1.87	2.60	3.86	3.88	5.25	3.90	5.46E-13	10.47	9.56	10.12	9.32	10.32	9.49	10.21	9.16
A1850133	AI850133	223864	Rap guanine nucleotide exchange factor (GEF) 3	Rapgef3	15	7.45E-01	-1.04	1.26E-02	-2.65	2.67E-01	-1.46	2.66E-02	-2.17	-1.83	0.18	1.53	0.62	1.32	0.91	3.61E-03	7.15	7.10	8.28	6.88	7.97	7.42	7.42	6.30
H4044F09	BQ557208					3.00E-01	-2.31	2.83E-01	-1.79	1.62E-02	-1.93	3.04E-01	-1.29	-1.83	0.57	0.60	1.46	0.57	0.80	2.76E-04	4.95	3.74	5.36	4.52	5.24	4.29	5.32	4.95
H4044G01	BQ557215	27360	Adducin 3 gamma	Add3	19	3.06E-01	-1.98	2.08E-01	-1.76	2.18E-01	-1.69	4.42E-03	-1.57	-1.81	0.57	0.71	0.69	1.81	0.94	1.18E-03	4.44	3.45	4.24	3.41	4.40	3.48	4.76	4.11
H4078E03	BQ562876	14017	Ecotropic viral integration site 2.	Evi2a	11	2.74E-02	-2.56	1.55E-01	-1.3	2.41E-02	-1.65	6.70E-03	-1.7	-1.80	1.32	0.80	1.35	1.70	1.29	2.52E-12	5.72	4.36	6.39	6.02	6.58	5.86	6.32	5.55
A1851309	AI851309	329867	Transmembrane 106c	Tmem106c	7	5.42E-02	-3.87	3.02E-01	-1.2	2.06E-02	-2.11	4.16E-01	-1.23	-1.88	1.12	1.42	0.64	0.45	0.91	4.57E-05	5.39	3.58	5.18	4.89	5.53	4.64	5.96	4.94
H3072D08	BG069071	67464	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	14	2.50E-04	-1.88	4.41E-06	-1.73	4.15E-06	-1.78	1.18E-07	-2.07	-1.87	2.60	3.86	3.88	5.25	3.90	5.46E-13	10.47	9.56	10.12	9.32	10.32	9.49	10.21	9.16
H4079G06	BQ563102	12843	Procollagen, type I, alpha.	Col1a2	6	8.58E-02	-2.42	9.24E-02	-1.58	6.21E-02	-1.65	1.51E-01	-1.48	-1.78	0.99	0.97	1.08	0.81	0.96	0.00E+000	4.95	3.67	4.84	4.28	4.87	4.14	4.55	3.88
A1851383	AI851383	215160	Rhomboid domain containing.	Rhbdd2	5	8.67E-03	-3.33	6.43E-04	-1.9	2.39E-03	-1.49	1.93E-05	-2.71	-1.78	1.83	2.34	2.02	3.37	2.34	3.24E-10	7.22	6.80	7.94	7.36	6.93	7.94	6.43	5.49
A1846370	AI846370	56307	Methionine aminopeptidase.	Metap2	10	1.17E-03	-2.36	2.66E-02	-1.48	6.32E-02	-1.27	5.19E-02	-1.59	-1.68	2.17	1.32	1.08	1.14	1.43	1.75E-08	6.88	6.64	6.54	5.97	6.88	6.53	6.36	5.69
H3066H07	BG061941	15369	Heme oxygenase (decycling).	Hmox2	16	3.33E-03	-2.31	1.57E-02	-1.58	2.75E-02	-1.3	2.53E-01	-1.43	-1.66	1.67	1.48	1.03	0.64	1.21	2.98E-07	6.09	4.88	5.80	5.13	5.86	5.48	5.52	5.30
H4064F12	BQ560584					4.91E-01	-1.36	3.62E-02	-2.04	1.19E-01	-1.46	8.84E-03	-1.73	-1.65	0.38	1.24	0.89	1.62	1.03	9.68E-05	4.98	4.54	5.29	4.25	4.47	3.92	5.20	4.41
H4019C12	BQ552979					4.91E-01	-1.36	3.62E-02	-2.04	1.19E-01	-1.46	8.84E-03	-1.73	-1.65	0.38	1.24	0.89	1.62	1.03	9.68E-05	4.98	4.54	5.29	4.25	4.47	3.92	5.20	4.41
A1844710	AI844710		Transcribed locus		14	2.75E-02	-1.46	2.61E-03	-1.59	1.08E-05	-1.87	4.00E-04	-1.66	-1.65	1.31	1.95	3.56	2.47	2.32	9.82E-07	7.53	6.98	7.40	6.73	7.57	6.67	7.24	6.50
H4015H03	BQ552441	73683	Developmental pluripotency associated	Dppa4	16	2.04E-02	-1.82	5.57E-02	-1.58	3.15E-01	-1.71	3.71E-01	-1.23	-1.63	0.40	0.72	0.74	0.49	0.84	2.14E-06	5.63	4.66	4.97	4.11	5.13	4.25	5.10	4.19
A1847221	AI847221		Transcribed locus		4	1.48E-01	-1.79	1.38E-01	-1.42	2.46E-02	-2.35	2.04E-01	-1.42	-1.61	0.56	0.73	1.35	0.71	0.84	1.42E-06	8.18	7.88	8.23	7.73	8.70	7.47	7.65	7.14
A1844240	AI844240		Transcribed locus		4	1.11E-01	-1.73	3.16E-02	-1.75	6.07E-02	-1.23	1.36E-01	-1.81	-1.63	0.91	1.28	1.09	0.85	1.03	0.00E+000	7.94	7.15	8.51	7.72	8.14	7.84	8.37	7.51
7216T7	AI668740		Transcribed locus		16	3.10E-01	-1.23	1.96E-01	-1.42	2.48E-02	-2.35	2.04E-01	-1.42	-1.61	0.56	0.73	1.35	0.71	0.84	1.42E-06	8.18	7.88	8.23	7.73	8.70	7.47	7.65	7.14
A1853831	AI853831	51788	H2A histone family, member J	H2afjz	6	5.97E-02	-2.37	4.32E-02	-1.54	1.62E-01	-1.08	1.43E-01	-1.42	-1.59	1.07	1.21	0.70	0.83	0.84	7.40E-03	6.80	5.96	6.49	5.90	6.07	5.96	6.48	5.97
H3126A07	AI853853					1.01E-01	-1.61	2.71E-02	-1.79	1.15E-01	-1.43	8.77E-02	-1.5	-1.58	0.94	1.38	0.90	0.98	1.05	0.00E+000	5.39	4.71	5.46	4.62	5.27	4.76	5.36	4.24
H4046E06	BQ557995	108148	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2	Galnt2	8	2.23E-01	-1.61	8.62E-02	-1.53	3.63E-01	-1.47	6.55E-02	-1.71	-1.58	0.68	0.99	0.50	1.07	0.81	8.04E-10	5.39	4.70	5.62	5.01	5.70	5.14	5.28	4.51
H4064D04	BQ560445	72722	RIKEN cDNA 2810405J04 gene	2810405J04Rik	17	1.16E-02	-1.84	6.54E-03	-1.88	1.93E-01	-1.34	4.13E-01	-1.23	-1.57	1.55	1.71	0.73	0.45	1.11	2.94E-04	5.95	5.08	8.74	4.84	5.68	5.26	5.65	5.35
H4077H04	BQ562771	21929	Tumor necrosis factor, alpha-induced protein	Tnfaip3	10	2.13E-01	-1.55	3.56E-01	-1.23	1.29E-01	-1.82	1.57E-01	-1.57	-1.57	0.70	0.51	0.87	1.20	0.82	2.39E-08	5.64	5.01	5.45	5.16	5.70	4.84	5.77	5.03
A1853572			Transcribed locus, strongly similar to XP_916099.1 similar to Cyclin I [Mus musculus]		11	1.56E-03	-1.59	2.86E-07	-1.82	1.40E-05	-1.63	1.38E-02	-1.24	-1.57	2.09	4.88	3.48	1.50	2.99	7.37E-05	8.04	7.36	7.93	7.07	8.05	7.34	7.40	7.08
H3106B12	CK334763	434459	Predicted gene, EG434459	EG434459	14	3.56E-02	-2.08	8.27E-02	-1.47	4.22E-02	-1.5	2.17E-01	-1.22	-1.57	1.24	1.00	1.20	0.69	1.03	2.22E-16	6.05	5.00	5.89	5.33	6.71	6.12	5.95	5.66
H4006C12	BQ550938	140919	Solute carrier family 7 (sodium-dependent amino acid transporter), member 6	Slc7a6	7	1.99E-03	-1.75	5.28E-02	-1.33	2.98E-02	-1.18	4.97E-05	-2.01	-1.57	2.03	1.13	1.29	3.10	1.89	2.53E-05	8.00	7.20	8.22	7.81	10.54	10.30	7.32	6.26
350I83	AI413229	101285	Expressed sequence AU067826	AU067826	6	2.04E-01	-1.59	9.19E-03	-1.86	3.21E-01	-1.39	2.00E-01	-1.42	-1.57	0.71	1.61	0.55	0.72	0.90	2.00E-04	5.83	5.16	5.26	4.37	4.88	4.40	5.34	4.84
H4022F10	BQ553535	26968	Immunoglobulin superfamily containing leucine-rich repeat	Islr	9	6.74E-01	-1.3	4.49E-02	-1.44	3.42E-02	-2.21	1.61E-01	-1.3	-1.56	0.23	1.18	1.25	0.79	0.86	2.29E-04	5.26	4.88	5.11	4.59	5.70	4.55	5.15	4.77
H4077F05	BQ562727	12487	CD28 antigen	Cd28	1	1.43E-01	-1.58	1.20E-02	-1.54	2.83E-01	-1.41	2.35E-01	-1.69	-1.56	0.83	1.54	0.60	0.59	0.89	7.13E-05	6.45	5.79	5.93	5.31	5.85	5.36	6.69	5.94
A1846695	AI846695	19317	Quaking	Qk	17	2.63E-02	-2.11	1.13E-01	-1.47	3.85E-01	-1.23	2.86E-01	-1.39	-1.55	1.33	0.90	0.48	0.59	0.83	1.32E-05	5.01	3.93	5.03	4.48	4.93	4.63	4.91	4.44
H3022A11	BG064830	109624	Calcineurin 1	Cald1	6	6.93E-01	-1.11	4.25E-01	-1.19	2.66E-02	-2.05	9.54E-03	-1.75	-1.53	0.22	0.44	1.32	1.60	0.92	7.70E-03	5.69	5.53	6.43	6.17	6.83	5.79	6.53	5.72
A1846612	AI846612	26921	Mitogen-activated protein kinase kinase kinase 4	Map4k4	1	1.64E-01	-1.8	1.96E-01	-1.35	2.39E-02	-1.45	3.22E-02	-1.48	-1.52	0.78	0.73	1.35	1.27	1.03	1.80E-10	6.43	5.58	5.37	4.94	6.05	5.52	6.74	5.78
H3037G01	BG065994					2.71E-01	-1.45	3.18E-01	-1.37	3.81E-01	-1.26	1.16E-02	-1.98	-1.52	0.61	0.55	0.48	1.55	0.80	1.48E-03	4.94	4.41	5.27	4.82	4.73	4.40	5.53	4.54
H4079D12	BQ563545	74155	Transcribed locus		11	4.56E-01	-1.3	1.32E-02	-1.61	1.24E-01	-1.5	9.88E-02	-1.59	-1.52	0.41	1.48	0.68	0.95	0.91	4.76E-05	5.58	5.20	5.50	4.76	5.46	4.87	6.02	5.35
H4064E12	BQ560460	58994	Sphingomyelin phosphodiesterase 3, neutr	Smpd3	8	5.56E-01	-1.21	3.14E-02	-1.87	1.22E-01	-1.53	2.22E-02	-1.43	-1.51	0.32	1.28	0.88	1.37	0.96	5.47E-05	6.35	6.07	6.78	5.88	6.68	6.07	6.16	5.65
H4076B07	BQ562466		NOD-derived CD11c+ dendritic cell cDNA, RIKEN full-length enriched library, clone:F630026I02 product:hypothetical protein, full insert sequence		13	4.90E-01	-1.22	1.84E-01	-1.43	5.28E-02	-1.61	9.26E-02	-1.76	-1.51	0.38	0.75	1.13	0.96	0.81	7.10E-07	4.63	4.34	5.09	4.57	4.92	4.23	5.37	4.55
A1841665	AI841665	17387	Matrix metalloproteinase 14 (membrane-inserted	Mmp14	14	1.37E-01	-1.23	5.55E-02	-1.48	4.71E-02	-2.23	7.20E-01	-1.06	-1.50	0.84	1.12	1.16	0.20	0.83	1.97E-04	5.78	5.48	6.09	5.52	6.37	5.21	5.94	5.86

Table 3.1: Genes expressed higher in B6J in all brain regions. CLID, clone ID; GBID, Genebank accession; Chr, chromosome; CoStr, cortex and striatum (orange); Hip, hippocampus (yellow); Vb, ventral brain region (green); Cer, Cerebellum (blue); *p*, *p*-value from t-test; FC, fold change; AVG, average; *d*, effect size; and (-), higher in B6J. Criterion for inclusion is: p, z-test<0.01; average |d|>0.8; and average |fold change|>1.5.

CLID	GBID	Entrez	Name	Symbol	Chr	CoStr p	CoStr FC	Hip p	Hip FC	Vb p	Vb Fc	Cer p	Cer FC	AVG FC	CoStr d	Hip d	Vb d	Cer d	AVG d	p_Ztest	J CoStr AVG	J Hip AVG	NCrl Hip AVG	J Vb AVG	NCrl Vb Avg	J Cer AVG	NCrl Cer AVG
643560	AA451673	56095	GTP binding protein 2	Gtpbp2	17	7.84E-02	2.17	3.39E-01	1.18	1.34E-01	1.29	5.45E-02	1.42	1.52	-1.02	-0.53	-0.85	-1.12	-0.88	1.03E-11	5.29	5.61	6.16	5.88	6.74	6.53	7.04
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	2.99E-01	1.45	5.81E-02	1.61	1.52	-0.49	-1.36	-0.58	-1.10	-0.81	3.98E-07	5.04	5.55	5.23	4.85	5.38	4.93	5.61
H4003A06	BQ550379	72305	RIKEN cDNA 2210021J22 gene	2210021J22Rik	15	5.93E-01	1.32	7.56E-02	1.69	5.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H4003F09	BQ554629	13805	Ensoglin	Eng	2	3.35E-01	1.48	1.08E-01	1.74	2.49E-02	1.75	4.70E-01	1.25	1.54	-0.26	-1.32	-2.46	-0.83	-1.26	6.10E-05	5.01	5.68	5.35	4.83	5.62	5.09	5.35
H4007E05						8.41E-01	1.08	1.45E-01	1.61	1.94E-01	1.54	1.45E-01	1.54	1.54	-0.08	-1.08	-1.29	-2.41	-1.22	1.46E-05	4.95	5.47	5.16	4.64	5.36	5.04	5.54
H4007A08	BQ550347	99922	Endothelial differentiation-related factor 1	Edf1	2	8.41E-01	1.1	7.45E-03	2.21	4.14E-02	1.35	1.94E-04	1.48	1.54	-0.11	-1.57	-1.20	-2.67	-1.41	7.94E-03	5.11	5.25	4.59	5.73	4.93	5.36	5.57	5.83
A1852009	AI852009	269 3	G protein-regulated inducer of neurite outgrowth 1	Gprin1	13	9.49E-02	1.51	2.94E-01	1.22	4.77E-02	1.85	1.09E-01	1.58	1.54	-0.96	-0.30	-1.16	-0.92	-0.91	8.22E-15	5.48	6.07	5.75	6.04	5.42	6.31	4.89	5.56
A1852010	AI852010	69368	WD repeat and FYVE domain containing 1	Wdfy1	1	9.23E-01	2.34	5.82E-01	1.05	1.34E-02	1.36	4.13E-03	1.42	1.55	-0.07	-0.13	-1.51	-1.60	-1.04	4.47E-04	4.85	5.18	5.24	4.80	5.60	5.00	5.46	5.67
H4007E05	BQ550347					5.24E-03	1.79	9.64E-04	1.77	4.41E-02	1.6	5.17E-01	1.13	1.57	-1.77	-2.22	-1.18	-0.36	-1.38	6.07E-04	5.19	6.03	5.36	6.18	5.35	6.03	5.70	5.88
H4008A07	BQ550379	72305	Ensoglin	Eng	2	3.35E-01	1.48	1.08E-01	1.74	2.49E-02	1.75	4.70E-01	1.25	1.54	-0.26	-1.32	-2.46	-0.83	-1.26	6.10E-05	5.01	5.68	5.35	4.83	5.62	5.09	5.35
H4007A08	BQ550347	99922	Endothelial differentiation-related factor 1	Edf1	2	8.41E-01	1.1	7.45E-03	2.21	4.14E-02	1.35	1.94E-04	1.48	1.54	-0.11	-1.57	-1.20	-2.67	-1.41	7.94E-03	5.11	5.25	4.59	5.73	4.93	5.36	5.57	5.83
A1852009	AI852009	269 3	G protein-regulated inducer of neurite outgrowth 1	Gprin1	13	9.49E-02	1.51	2.94E-01	1.22	4.77E-02	1.85	1.09E-01	1.58	1.54	-0.96	-0.30	-1.16	-0.92	-0.91	8.22E-15	5.48	6.07	5.75	6.04	5.42	6.31	4.89	5.56
A1852010	AI852010	69368	WD repeat and FYVE domain containing 1	Wdfy1	1	9.23E-01	2.34	5.82E-01	1.05	1.34E-02	1.36	4.13E-03	1.42	1.55	-0.07	-0.13	-1.51	-1.60	-1.04	4.47E-04	4.85	5.18	5.24	4.80	5.60	5.00	5.46	5.67
H4007E05	BQ550347					5.24E-03	1.79	9.64E-04	1.77	4.41E-02	1.6	5.17E-01	1.13	1.57	-1.77	-2.22	-1.18	-0.36	-1.38	6.07E-04	5.19	6.03	5.36	6.18	5.35	6.03	5.70	5.88
H4008A07	BQ550379	72305	Ensoglin	Eng	2	3.35E-01	1.48	1.08E-01	1.74	2.49E-02	1.75	4.70E-01	1.25	1.54	-0.26	-1.32	-2.46	-0.83	-1.26	6.10E-05	5.01	5.68	5.35	4.83	5.62	5.09	5.35
H4007A08	BQ550347	99922	Endothelial differentiation-related factor 1	Edf1	2	8.41E-01	1.1	7.45E-03	2.21	4.14E-02	1.35	1.94E-04	1.48	1.54	-0.11	-1.57	-1.20	-2.67	-1.41	7.94E-03	5.11	5.25	4.59	5.73	4.93	5.36	5.57	5.83
A1852009	AI852009	269 3	G protein-regulated inducer of neurite outgrowth 1	Gprin1	13	9.49E-02	1.51	2.94E-01	1.22	4.77E-02	1.85	1.09E-01	1.58	1.54	-0.96	-0.30	-1.16	-0.92	-0.91	8.22E-15	5.48	6.07	5.75	6.04	5.42	6.31	4.89	5.56
A1852010	AI852010	69368	WD repeat and FYVE domain containing 1	Wdfy1	1	9.23E-01	2.34	5.82E-01	1.05	1.34E-02	1.36	4.13E-03	1.42	1.55	-0.07	-0.13	-1.51	-1.60	-1.04	4.47E-04	4.85	5.18	5.24	4.80	5.60	5.00	5.46	5.67
H4007E05	BQ550347					5.24E-03	1.79	9.64E-04	1.77	4.41E-02	1.6	5.17E-01	1.13	1.57	-1.77	-2.22	-1.18	-0.36	-1.38	6.07E-04	5.19	6.03	5.36	6.18	5.35	6.03	5.70	5.88
429422	AA429884	19824	Tripartite motif protein 19	Trim19	17	1.46E-01	1.42	9.79E-02	1.42	2.97E-02	2.05	2.20E-01	1.45	1.59	-0.52	-0.95	-2.29	-0.71	-0.94	7.24E-14	4.94	5.45	4.79	5.29	4.28	5.32	4.71	5.25
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H4003A06	BQ550379	72305	RIKEN cDNA 2210021J22 gene	2210021J22Rik	15	5.93E-01	1.32	7.56E-02	1.69	5.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H4008A07	BQ550379	72305	Ensoglin	Eng	2	3.35E-01	1.48	1.08E-01	1.74	2.49E-02	1.75	4.70E-01	1.25	1.54	-0.26	-1.32	-2.46	-0.83	-1.26	6.10E-05	5.01	5.68	5.35	4.83	5.62	5.09	5.35
H4007A08	BQ550347	99922	Endothelial differentiation-related factor 1	Edf1	2	8.41E-01	1.1	7.45E-03	2.21	4.14E-02	1.35	1.94E-04	1.48	1.54	-0.11	-1.57	-1.20	-2.67	-1.41	7.94E-03	5.11	5.25	4.59	5.73	4.93	5.36	5.57	5.83
A1852009	AI852009	269 3	G protein-regulated inducer of neurite outgrowth 1	Gprin1	13	9.49E-02	1.51	2.94E-01	1.22	4.77E-02	1.85	1.09E-01	1.58	1.54	-0.96	-0.30	-1.16	-0.92	-0.91	8.22E-15	5.48	6.07	5.75	6.04	5.42	6.31	4.89	5.56
H4007E05	BQ550347					5.24E-03	1.79	9.64E-04	1.77	4.41E-02	1.6	5.17E-01	1.13	1.57	-1.77	-2.22	-1.18	-0.36	-1.38	6.07E-04	5.19	6.03	5.36	6.18	5.35	6.03	5.70	5.88
A1852010	AI852010	69368	WD repeat and FYVE domain containing 1	Wdfy1	1	9.23E-01	2.34	5.82E-01	1.05	1.34E-02	1.36	4.13E-03	1.42	1.55	-0.07	-0.13	-1.51	-1.60	-1.04	4.47E-04	4.85	5.18	5.24	4.80	5.60	5.00	5.46	5.67
429422	AA429884	19824	Tripartite motif protein 19	Trim19	17	1.46E-01	1.42	9.79E-02	1.42	2.97E-02	2.05	2.20E-01	1.45	1.59	-0.52	-0.95	-2.29	-0.71	-0.94	7.24E-14	4.94	5.45	4.79	5.29	4.28	5.32	4.71	5.25
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H3058010	BG087643					3.70E-01	1.42	6.68E-02	1.59	6.97E-02	1.67	9.23E-02	1.46	1.54	-0.29	-1.33	-1.11	-0.97	-0.85	5.95E-06	4.56	4.95	4.78	5.53	3.95	4.69	4.16	4.70
H0048G07	BQ067030	64292	Prostaglandin E synthase	Ptges	2	1.44E-01	1.64	3.19E-01	1.34	1.99E-01	1.79	5.18E-02	1.59	1.59	-0.33	-0.25	-0.78	-1.14	-0.82	6.31E-12	4.03	4.74	4.86	5.38	3.37	4.21	3.92	4.59
H3018H02	BG077462	67967	Polymerase (DNA-directed), delta 3, accessory subunit	Pold3	7	1.44E-01	1.48	2.65E-02	1.97	2.67E-01	1.32	1.57E-01	1.6	1.59	-0.83	-1.13	-0.62	-0.80	-0.89	3.72E-09	5.39	5.96	4.95	5.92	4.79	5.19	5.35	6.03
H3067C06						3.79E-02	2.39	1.03E-01	1.48	1.09E-01	1.25	2.24E-01	1.29	1.60	-0.96	-0.68	-0.96	-0.96	-0.96	2.06E-13	6.00	7.26	6.91	6.69	7.02	7.13	7.50
H3064H01	BG086380					7.51E-01	1.81	1.64E-01	1.91	1.34E-01	1.49	1.60	-0.17	-0.78	-0.26	-0.90	-0.60	-0.03	-0.58	6.90E-03	4.74	5.00	4.91	5.77	3.63	4.96	4.81	5.38
635673	AA451264	74157	RIKEN cDNA 1300018I05 gene	1300018I05Rik	17	2.48E-01	1.26	2.40E-02	2.78	2.48E-02	1.33	4.58E-01	1.08	1.61	-0.54	-1.35	-1.34	-0.41	-0.34	1.11E-04	8.77	9.10	8.75	8.17	8.58	8.07	8.19
UTDHES1712807						7.16E-02	1.69	2.76E-01	1.32	6.22E-02	1.35	1.05E-01	1.72	1.62	-1.04	-0.93	-1.46	-0.81	-0.98	2.00E-15	4.88	5.64	4.49	4.89	4.76	5.19	5.11	6.19
H3021H12	AA449233	74318	Homeobox only domain	Hod	5	7.16E-02	1.69	2.76E-01	1.32	6.22E-02	1.35	1.05E-01	1.72	1.62	-1.04	-0.93	-1.46	-0.81	-0.98	2.00E-15	4.88	5.64	4.49	4.89	4.76	5.19	5.11	6.19
H3021H12	BQ079233	110583	PR domain containing 2, with ZNF domain	Prdm2	4	2.39E-01	1.46	5.11E-01	1.06	1.23E-02	2.15	1.44E-02	1.68	1.74	-0.68	-0.26	-1.55	-0.49	-0.84	8.37E-03	5.09	5.04	4.64	4.85	5.28	4.64	4.96	5.12
H3154H02	BG076037	101540	Protein kinase D2	Prkd2	7	4.29E-01	1.61	6.48E-03	2.35	4.27E-02	1.68	7.74E-02	1.41	1.76	-0.44	-1.71	-1.19	-1.02	-1.09	3.33E-05	4.49	5.17	4.31	5.54	4.32	5.07	4.63	5.13
A1846618	AI846618					1.24E-01	2.35	2.26E-01	1.38	3.18E-01	1.43	2.43E-02	1.91	1.77	-0.89	-0.38	-0.56	-1.35	-0.87	6.27E-07	3.81	5.05	4.73	5.19	4.22	4.73	3.86	4.79
A1846618	AI846618					1.24E-01	2.35	2.26E-01	1.38	3.18E-01	1.43	2.43E-02	1.91	1.77	-0.89	-0.38	-0.56	-1.35	-0.87	6.27E-07	3.81	5.05	4.73	5.19	4.22	4.73	3.86	4.79
A1846790	AI846790					2.41E-01	1.56	5.88E-01	1.19	7.40E-02	2.3	2.61E-02	2.08	1.78	-0.65	-0.30	-1.03	-1.31	-0.82	1.95E-04	4.90	5.54	4.95	5.30	4.49	5.70	4.89	6.04
H4003D06	BQ554441	226 2	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1	Yes1	5	3.30E-03	2.36	9.40E-01	1.02	3.01E-02	2.06	2.26E-02	1.75	1.80	-1.89	-0.34	-1.29	-1.37	-1.15	3.42E-03	4.56	5.61	4.76	4.79	4.10	5.14	4.53	5.13
H4001A10	BQ550551	19264	Polovirus receptor-related 2	Pvrl2	7	7.50E-02	1.56	4.20E-03	2.4	2.48E-02	2.16	6.53E-01	1.1	1.81	-1.03	-1.52	-1.34	-0.26	-1.11	6.81E-04	5.26	5.91	4.49	5.75	4.20	5.31	5.09	5.22
H4001G11	BQ550577	100653	Reprimo-like 1 mRNA, complete cds	Rtf	4	4.64E-01	1.78	1.78E-01	1.73	1.78E-01	1.77	1.09E-01	1.81	1.77	-0.37	-0.71	-0.84	-0.71	-0.77	1.45E-04	4.89	5.04	4.84	5.57	4.10	5.46	4.78	5.13
A1853705	AI853705		Transcribed locus			5.24E-01	1.83	7.35E-02	2.09	4.30E-03	2.19	1.77E-01	1.17	1.82	-0.08	-1.03	-1.82	-0.77	-1.08	3.09E-05	5.31	6.18	5.25	6.32	6.10	7.23	10.94	11.17
476947	AA427464	67383	RIKEN cDNA 2410127L17 gene	2410127L17Rik	9	4.81E-01	1.49	3.32E-02	1.24	1.72E-01	1.86	5.21E-02	2.73	1.86	-0.38	-1.36	-0.77	-1.13	-0.89	7.83E-06	3.70	4.27	4.01	3.06	4.03	3.36	4.81
477144	AA426170	16070	Insulin-like growth factor binding protein 4	Igfbp4	11	1.97E-01	1.72	1.48E-03	2.76	2.00E-01	1.96	4.26E-01	1.51	1.86	-0.72	-2.11	-0.72	-0.41	-0.95	9.04E-03	3.91	4.18	3.54	5.00	3.13	4.10	3.68	4.27
R67870	AA464402	380538	Ras homolog gene family, member A	Rhoa	X	2.07E-01	1.79	3.53E-03	2.41	2.58E-02	2.11	2.39E-02	1.96	2.02	-0.67	-1.53	-1.54	-0.95	-1.05	1.39E-05	3.82	4.67	4.77	5.41	3.58	4.65	4.10	4.57
483572	AA427658	226388	Solute carrier family 35, member C1	Slc35c1	2	2.14E-01	1.38	1.10E-01	2.4	2.04E-02	1.99	3.39E-01	1.69	1.87	-0.70	-0.31	-1.40	-0.53	-0.98	2.49E-06	5.44	5.90	5.18	6.44	4.44	5.44	4.86	5.62
H3020B12	BG064476	58180	Hydroxysteroid (17-beta) dehydrogenase 12	Hsd17b12	16	1.64E-02	1.87	3.10E-02	2.29	5.09E-02	2.29	6.54E-01	1.09	1.89	-1.46	-1.28	-1.14	-0.24	-1.03	1.14E-04	4.49	5.40	4.50	5.70	4.68	5.88	4.88	5.00
H3012A12	BG063730	239569	Alpha-1,4-galactosyltransferase	A4galt	15	3.80E-01	2.13	4.79E-02	2.19	2.43E-01	1.82	1.09E-01	2.24	1.93	-0.39	-1.16	-0.69	-0.79	-0.91	1.78E-11	4.49	5.36	4.17	5.10	4.64	5.15	5.17	5.93
H3021A03	BG063788					1.64E-02	1.87	3.10E-02	2.29	5.09E-02	2.29	6.54E-01	1.09	1.89	-1.46	-1.28	-1.14	-0.24	-1.03	1.14E-04	4.49	5.40	4.50	5.70	4.68	5.88	4.88	5.00
523447	AA427847	209586	NudC domain containing 3	Nudcd3	11	1.41E-01	1.89	2.64E-02	3.27	9.58E-01	1.02	3.48E-02	1.68	1.97	-0.83	-1.13	-0.03	-1.25	-0.98	3.95E-03	4.59	5.51	4.77	5.48	5.03	5.05	5.33	6.09
H3042A04						6.59E-02	2.02	2.28E-01	2.47	4.11E-01	1.3	1.42E-02	2.16	1.99	-1.07	-0.17	-0.45	-1.50	-0.92	5.01E-05	4.30	5.31	4.27	5.58	4.04	4.42	4.38	5.48
A1842091	AA426091	243009	Stanniocalcin 2	Stc2	5	1.99E-01	2.64	1.48E-02	2.83	1.76E-01	1.85	5.42E-01	1.25	2.01	-0.93	-1.49	-0.76	-0.24	-0.89	3.94E-05	3.81	5.25	4.35	4.76	4.10	4.61	4.41	4.72
A1842091	AI842091	243009	Stanniocalcin 2	Stc2	5	1.99E-01	2.64	1.48E-02	2.83	1.76E-01	1.85	5.42E-01	1.25	2.01	-0.93	-1.49	-0.76	-0.24	-0.89	3.94E-05	3.81	5.25	4.35	4.76	4.10	4.61	4.41	4.72
H3152H02	BG075873	67721	TBC1 domain family, member 20	Tbc1d20	2	1.10E-02	2.54	4.14E-01	1.48	4.99E-01	1.24	8.28E-02	2.99	2.06	-1.57	-0.45	-0.37	-1.00	-0.85	2.29E-03	4.56	5.91	3.55	4.72	4.27	4.57	4.61	6.19
H4048G05	BQ557869	114713	RAS p21 protein activator 2	Rasa2	9	8.91E-01	1.04	3.11E-02	3.05	5.39E-02	2.79	2.88E-02	1.42	2.08	-0.07	-1.28	-1.10	-1.30	-0.94	1.23E-03	4.76	4.82	4.71	4.59	6.07	4.74	5.24
A1842458	AI842458	243009	Stanniocalcin 2	Stc2	5	1.99E-01	2.64	1.48E-02	2.83	1.76E-01	1.85	5.42E-01	1.25	2.01	-0.93	-1.49	-0.76	-0.24	-0.89	3.94E-05	3.81	5.25	4.35	4.76	4.10	4.61	4.41	4.72
H4033E01	BQ558328	69368	WD repeat and FYVE domain containing 1	Wdfy1	1	9.63E-03	1.81	4.18E-03	2.22	1.43E-06	2.73	4.68E-02	1.75	2.13	-1.60	-1.33	-4.26	-1.17	-2.21	1.47E-03	6.57	7.43	6.17	7.52	6.50	7.94	7.71
H3154D11	BG074383					3.86E-02	2.77	2.46E-01	1.27	8.18E-01	1.18	1.38E-02	3.51	2.18	-1.22	-0.35	-0.27	-1.50	-0.91	1.03E-03	3.50	4.97	5.04	5.38	5.12	5.36	4.17	5.86
S81755	AA451985	102826	Mitogen-activated protein kinase-activated protein kinase 3	Mapkapk3	9	1.12E-02	3.72	2.61E-01	1.69	5.27E-02	2.04	1.73E-01	1.36	2.20	-1.56	-0.53	-1.13	-0.77	-1.02	9.05E-07	2.86	4.76	4.58	2.87	3.89	3.64	4.09
H4034A01	BQ554420		Transcribed locus, strongly similar to NP_083496.2 repeat domain 17 [Mus musculus]		8	6.43E-01	1.11	1.03E-01	2.79	3.90E-02	2.74	5.13E-02	2.94	2.40	-0.25	-0.93	-1.25	-1.14	-0.89	6.22E-05	3.30	3.46	3.98	5.05	2.54	3.99	3.14	4.69
A1852640	AI852640	665672	RIKEN cDNA C030014I07 gene	C030014I07Rik	12	1.94E-01	2.4	7.87E-02	2.95	2.21E-01	2.17	4.74E-02	2.16	2.48	-0.75	-1.01	-0.68	-1.16	-0.90	8.88E-16	4.64	5.90	4.34	5.71	3.97	5.09	4.40	5.87
A1852640	AI852640	217302	Rich/Rho GTPase 2	Rhotekin2	12	1.94E-01	2.4	7.87E-02	2.95	2.21E-01	2.17	4.74E-02	2.16	2.48	-0.75	-1.01	-0.68	-1.16	-0.90	8.88E-16	4.64	5.90	4.34	5.71	3.97	5.09	4.40	5.87
H4047G10	BQ557713					4.67E-02	3.87	8.83E-02	2.62	1.29E-01	2.11	3.84E-01	1.78	2.60	-0.97	-0.48	-0.87	-1.03E-09	3.80	5.75	3.83	5.72	4.20	5.27	4.75	5.56
A1839860	AI839860	270076	Glutaryl-Coenzyme A dehydrogenase	Gcdh	8	5.24E-01	1.52	5.13E-03	3.16	2.18E-01	2	1.51E-03	4.2	2.72	-0.35	-1.14	-0.69	-2.10	-1.07	4.98E-03	3.09	3.69	2.89	4.54	2.54	3.04	2.70	4.77
H3044H01	BQ557035	237600	Tubulin tyrosine ligase-like family, member 6	Ttll6	11	8.80E-02	1.93	2.28E-01	2.27	4.98E-02	1.88	1.29E-01	3.09	2.06	-1.13	-0.38	-1.15	-0.20	-0.99	7.17E-03	2.10	4.31	2.89	4.18	2.89	3.48	2.23	3.35
A1842458	AI842458	21135	Deoxycytidine kinase	Dck1	3	3.95E-02	1.43	2.18E-01	1.23	3.22E-02	1.93	1.29E-01	2.22	1.79	-1.10	-0.55	-1.27	-1.09	-1.16	2.27E-04	5.22	6.03	5.84	6.28	6.28	6.53	7.36

Table 3.2: Genes expressed higher in B6C in all brain regions. CLID, clone ID; GBID, Genebank accession; Chr, chromosome; CoStr, cortex and striatum (orange); Hip, hippocampus (yellow); Vb, ventral brain region (green); Cer, Cerebellum (blue); *p*, *p*-value from t-test; FC, fold change; AVG, average; *d*, effect size; and (+), higher in B6C. Criterion for inclusion is: p, z-test<0.01; average |d|>0.8; and average |fold change|>1.5.

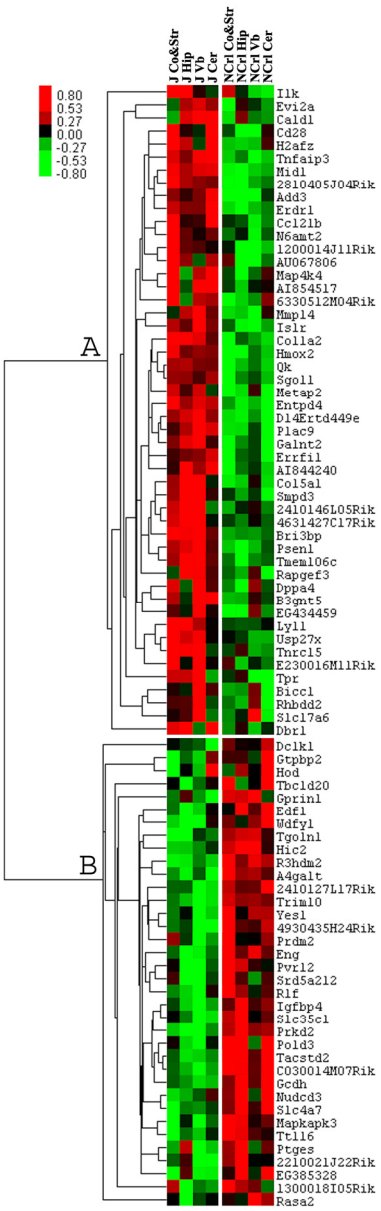


Figure 3.3: Genes significantly divergent between substrains Across all Four Brain Regions. Average standardized green channel intensities for each brain region were computed for each substrain and values were then centered, normalized and clustered (see methods). Red indicates increased expression and green indicates decreased expression in B6J relative to B6C. Gene symbols are shown to the right. Genes with large changes in the same direction across all four brain regions cluster into two groups (A and B) based on increased expression in B6J and decreased expression in B6J compared to B6C respectively. Criterion for selection was $p<0.01$, z-test; average $|d|>0.8$ and average |fold change|≥ 1.5. Co&Str, cortex and striatum; Hip, hippocampus; Vb, ventral brain region; and Cer, cerebellum.

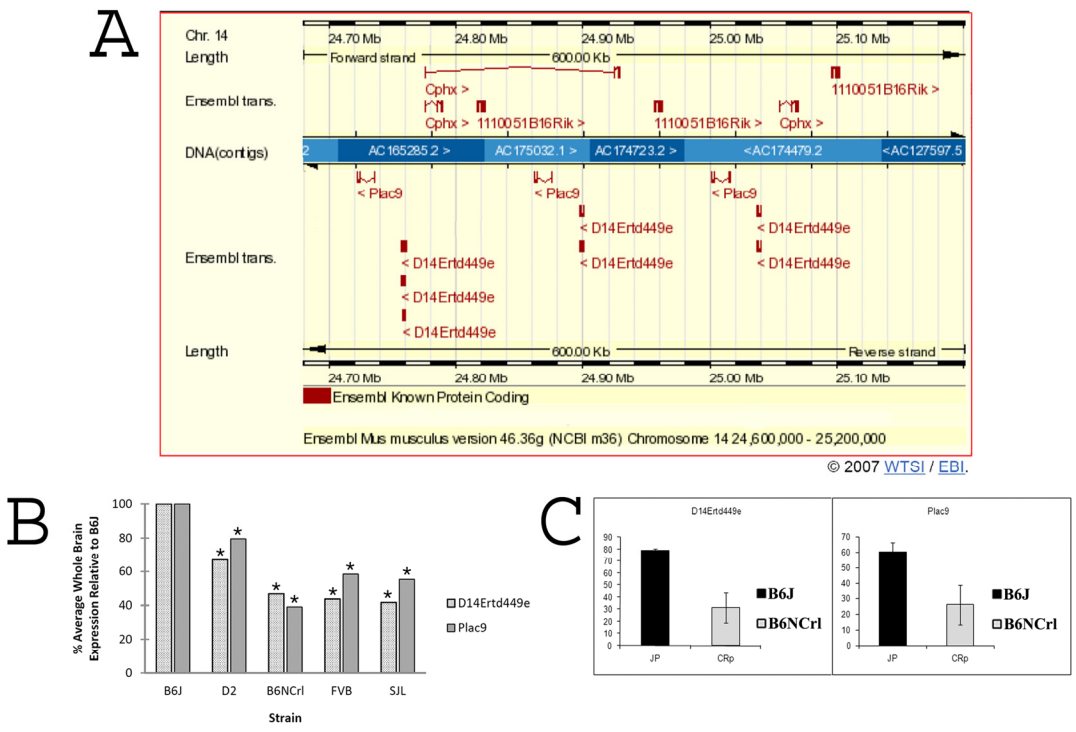


Figure 3.4: Putative changes in gene copy number between the B6 substrains on chromosome 14. (A) View of the chromosome region using the Ensemble Genome Browser, *Mus musculus* version 36 (NCBI). Trans.,transcript. *Cphx*, *1110051B16Rik, Plac9* and *D14Ertd449e* appear to be exist in multiple copies for the B6J reference strain. Only *Plac9* and *D14Ertd449e* are represented on our cDNA microarrays. (B) Expression data from other inbred strains. Probe set mean, SEM and n for *Plac9* (Affymetrix probe set 1452590_a_at) and *D14Ertd449e (*Affymetrix probe set 1428738_a_at) was gathered from WebQTL for B6J, DBA/2J (D2), FVB/NJ (FVB) and SJL/J (SJL) and tested for significant differences in expression by t-test. B6C (B6NCrl) expression was averaged across the brain regions for *Plac9* and *D14Ertd449e* and compared to the averaged brain region expression of B6J. Since expression is highest in the B6J strain it is shown as 100% and percent change from B6J is shown for all other strains. Asterisks indicate significant differences in mean expression between strains relative to the B6J strain ($p<0.001$, t-test). (C) Quantitative real time PCR validation. Representative probe values normalized against the endogenous control 18s rRNA. Since B6J is the substrain with the highest expression it is shown as 100% and the percent change from B6J was calculated for B6C from the qPCR data ($p=0.03$ and $p=0.005$ for Plac9 and D14Ertd449e respectively by two-tailed t-test).

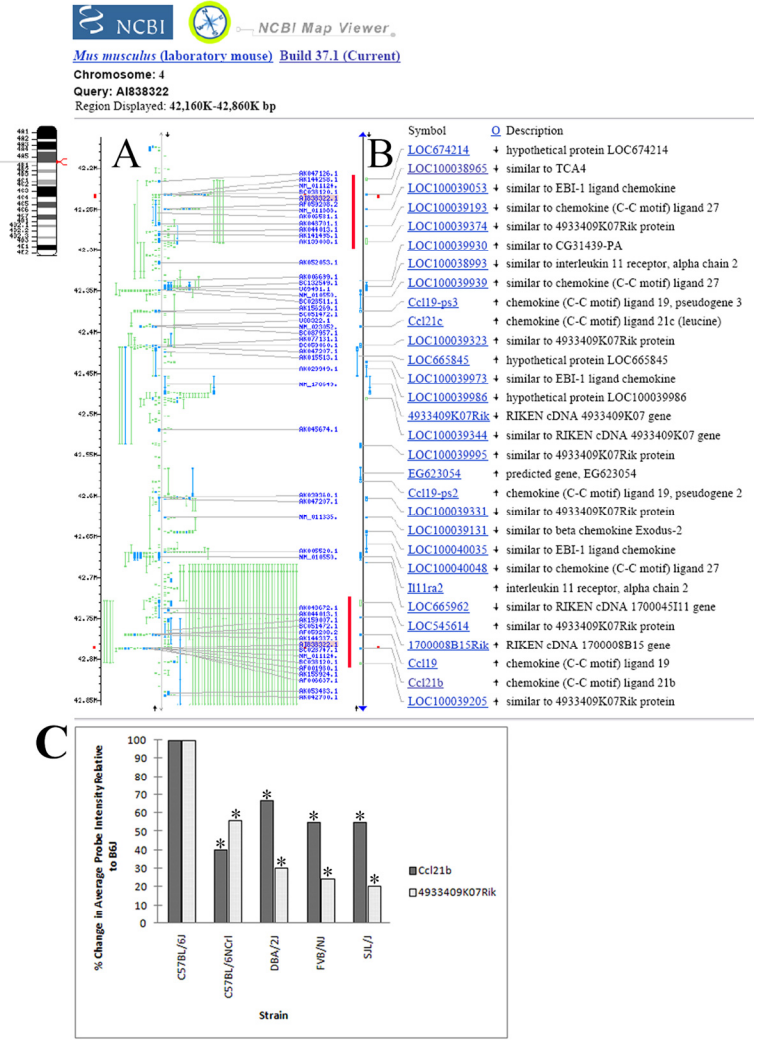


Figure 3.5 Putative chromosome 4 region containing copy number differences between substrains. (A) Known expressed transcripts within the chromosome 4 region. (B) Gene symbols, orientation of the gene and gene name. A and B were generated using NCBI Map Viewer Mus musculus build 37. Red square and red bar indicate duplicated locations for the transcript from our microarray data that represents Ccl21b. The cluster of 5 genes in this region appear to duplicated and are shown at the top and bottom of the top figure. Interestingly, it appears that they are opposite in orientation. (C) Analysis of expression across strains. This analysis was identical to the one for the chromosome 14 region with the exception that Ccl21b and 4933409K07Rik were represented by Affymetrix probe sets 1445238_at and 1447939_a_at respectively. Asterisks indicate significant deviation of strain means from the B6J strain mean ($p<0.001$).

Ch.	Interval (Mb)	Strain	LOD	Gene Symbols
1	174.0-175.8	minus	3.94	Pex19, Pea15a, Casq1, Atp1a2, Kcnj9, Darc, Cadm3, Ifi203
2	33.0-35.1	plus	3.93	Angptl2, Zbtb34, Mapkap1, Hspa5, Rabepk, Rab14, Gsn
13	54.4-55.5	minus	4.77	Cplx2, 4833439L19Rik, Higd2a, Cltb, Gprin1, Sncb, Unc5a, Zfp346, Rab24, Rgs14, Pdlim7
X	131.4-134.6	plus	3.93	Bex2, Tceal8, Bex1, Wbp5, Morf4l2, Plp1, Zcchc18, Mum1l1

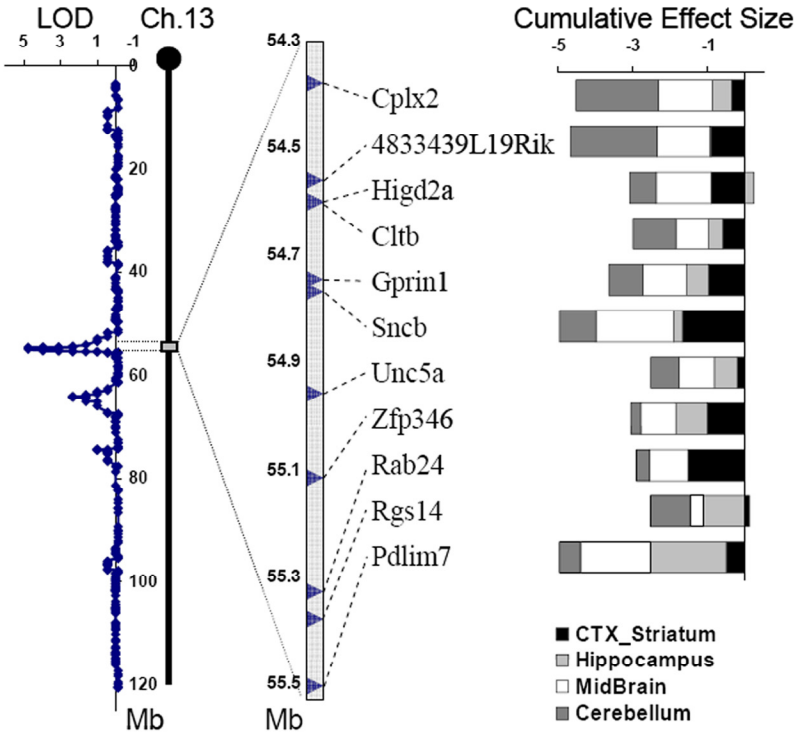


Figure 3.6: Regional chromosomal overrepresentation for significantly regulated transcripts. Four regions contain a greater number of co-localized and co-regulated genes across all brain regions and between strains than expected by chance. Ch, chromosome; interval, chromosomal region contain a significant number of regulated genes; minus designates expression that is lower in B6C and plus designates expression that is higher in B6J relative to B6C; LOD, logarithm of odds. By convention a LOD score greater than 3 indicates significance. Cumulative effect sizes for each gene are stacked and not overlayed.

Evidence for substrain differences in brain regional transcriptomes

At the 5% false discovery rate (q<0.05), microarray expression analysis detected 12, 13, 51 and 43 non-redundant transcripts from the cortex and striatum, hippocampus, ventral brain region and cerebellum respectively, that were highly significantly divergent between substrains in one or more of the brain regions (tables 3.3, 3.4 and 3.5). Some of these genes were identified by the previously described meta-analytic approach to be divergent between substrains in all four brain regions. The ventral brain region and the cerebellum contained the majority of the strain expression differences between the two B6 substrains and accounted for 43% and 36% of the total expression differences respectively.

Many genes known to be expressed in the brain and known to play a role in neurological processes were identified in each brain region at the highly stringent 5% false discovery rate. In the cortex and striatum, growth associated protein 43 (*Gap43*), adenosine deaminase, RNA-specific (*Adar*) and activating transcription factor 4 (Atf4) were expressed higher in B6J mice. Polymerase (DNA directed), eta (RAD 30 related) (*Polh*) was expressed 3.91 fold higher in the cortex and striatum of B6NCrl mice. The serotonin transporter (*Slc6a4*) was found to be expressed 14 fold higher in the hippocampus of B6J animals. The ventral brain region contained significant genes that played a role in hormone processing (proprotein convertase subtilisin/kexin type 2), intracellular trafficking (kinesin family members 5a and 5b), potassium channel activity (calsenilin, potassium voltage-gated channel, subfamily H (eag-related), member 3 and potassium large conductance calcium-activated channel, subfamily M, beta member 4), specific transcriptional regulation (Necdin) and general regulation of transcription and splicing at the C-terminal domain (CTD) of RNA polymerase II (Necdin, Cyclin dependent kinase 8, hexamethylene bis-acetamide inducible 1 and splicing factor,

arginine/serine-rich 3). In the cerebellum, the vesicular glutamate transporter (*Slc17a6*) was expressed 2 fold higher in B6J mice. Clusterin (*Clu*) was also expressed higher in B6J mice. Calcium/calmodulin-dependent protein kinase II gamma (*Camk2g*), Calcium channel, voltage-dependent, beta 3 subunit (*Cacnb3*), and a voltage-gated potassium channel beta subunit, shaker-related subfamily (*Kcnab3*) were expressed higher in B6NCrl cerebellum. Gene Ontology (GO) functional category overrepresentation analysis confirmed that potassium channel activity, microtubule-based processes and specific transcriptional repressor activity were significantly ($p<0.01$) overrepresented in the ventral midbrain and that calcium ion transport was significantly overrepresented in the cerebellum (table 3.6).

The Allen Brain Atlas (http://www.brainatlas.org/aba/) has B6J whole brain *in situ* data available for many genes and we used their database to verify some of our B6J expression levels and to determine if our transcripts were independently found to be expressed in the brain region for which they were detected. In general, the level of expression measured by our microarray analysis in B6J mice visually correlated well with the available *in situ* information from the Allen Brain Atlas. Although it should be stated that most of the transcripts had much broader expression in the brain and were not restricted to the region in which an expression difference was detected. It is likely that the low false discovery rate expression changes observed in each brain region resulted from a combination of the downstream effects of a polymorphism between substrains and/or local differences in brain gene regulation. The stringency of this analysis also limited detection to only the most significantly divergent transcripts. No *in situ's* were available nor were they completed for the B6C mice.

In silico transcription factor binding site (TFBS) overrepresentation analysis was again applied to the genes with similar patterns of expression in each substrain. In this

case, co-regulated transcripts from each brain region were independently screened using oPOSUM in order to locate a common transcription factor or TFBS in each brain region. Again, we were not able to identify a transcriptional regulator that could account for the majority of the observed changes between substrains in each region.

Cortex and Striatum

CLID	GBID	Name	Symbol	Entrez	Chr	Multiple Regions	Co&Str P	Co&Str Q	Co&Str FC	B6J AVG	B6NCrl AVG	Strain
H3119F05	BG073178	Midline 1	Mid1	17318	X	XVC	1.63E-07	0.00	-4.40	7.99	5.86	J
AI839337	AI839337					XHVC	2.20E-08	0.00	-2.63	8.55	7.16	J
H3108G07	BG072279	DNA segment, Chr 14, ERATO Doi 449, expressed	D14Ertd449e	66039	14	XVC	1.85E-07	0.00	-2.23	9.81	8.65	J
H4037F04	BQ556023	RIKEN cDNA 1110039B18 gene	1110039B18Rik	68796	5		1.52E-06	0.01	-1.93	5.94	4.99	J
H3157F06	BG076269	Growth associated protein 43	Gap43	14432	16		5.53E-06	0.02	-1.78	8.69	7.86	J
H3123A09	BG073479	Elastin microfibril interfacer 1	Emilin1	100952	5		5.07E-06	0.02	-1.72	6.37	5.58	J
AI842795	AI842795	Adenosine deaminase, RNA-specific	Adar	56417	3		2.27E-06	0.01	-1.52	8.73	8.13	J
H3124E06	BG073595	Activating transcription factor 4	Atf4	11911	15		9.36E-06	0.03	-1.37	9.24	8.78	J
H4037B11	BQ555948	Nudix (nucleotide diphosphate linked moiety X)-type motif 3	Nudt3	56409	17		9.82E-06	0.03	-1.33	12.22	11.80	J
H3074B02		RIKEN cDNA C030045D06 gene	C030045D06Rik	109294	1		1.14E-05	0.03	1.42	14.92	15.42	NCrl
AI844116	AI844116	Peroxisomal biogenesis factor 3	Pex3	56535	10		8.95E-06	0.03	1.62	6.95	7.65	NCrl
H4069E03	BQ561301	Polymerase (DNA directed), eta (RAD 30 related)	Polh	80905	17		1.03E-05	0.03	3.91	4.59	6.56	NCrl

Hippocampus

CLID	GBID	Name	Symbol	Entrez	Chr	Multiple Regions	Hip P	Hip Q	Hip FC	B6J AVG	B6NCrl AVG	Strain
AI846650	AI846650	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	Slc6a4	15567	11		1.21E-05	0.03	-14.36	4.80	0.96	J
AI839337	AI839337					XHVC	1.12E-08	0.00	-2.78	8.67	7.19	J
AI846216	AI846216	Zinc finger, CCHC domain containing 12	Zcchc12	72693	X		6.22E-06	0.02	-1.97	8.98	8.00	J
AI853572	AI853572	Transcribed locus, strongly similar to XP_916059.1 similar to Cyclin I [Mus musculus]			11	HV	2.86E-07	0.00	-1.82	7.93	7.07	J
H3068F10							7.89E-08	0.00	-1.76	8.52	7.71	J
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	67464	14	HVC	4.41E-06	0.02	-1.73	10.12	9.32	J
AI850284	AI850284	Activating transcription factor 4	Atf4	11911	15		9.36E-08	0.00	-1.44	10.75	10.22	J
572834	AI327284	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 4	Ndufb4	68194	16		2.08E-06	0.01	-1.19	12.13	11.88	J
H3086D03	BG070347	Transcribed locus			13		8.46E-06	0.03	1.28	9.30	9.66	NCrl
H3064G02	BG068369						3.10E-06	0.01	1.31	8.94	9.34	NCrl
H3058G04	BG067838	Suppressor of cytokine signaling 7	Socs7	192157	11		1.64E-05	0.04	1.63	7.70	8.40	NCrl
AI852992	AI852992	MiRNA containing gene	Mirg	373070	12		1.52E-05	0.04	1.91	6.91	7.85	NCrl
H3032F05						HV	2.01E-06	0.01	2.03	6.07	7.08	NCrl

Table 3.3: Transcripts highly significantly divergent between substrains in the cortex and striatum and hippocampus. A small number of genes are highly statistically significantly divergent between B6 substrains based on the criterion $Q<0.05$. (CLID, clone ID; GBID, GeneBank Acession; Entrez, Entrez ID; Chr, chromosome; X, cortex and striatum; H, Hippocampus; V, Ventral brain region; C, cerebellum.) Top panel is cortex and striatum and bottom panel is hippocampus.

Ventral Brain Region

CLID	GBID	Name	Symbol	Entrez	Chr	Multiple Regions	Vb P	Vb Q	Vb FC	B6J AVG	B6NCrl AVG	Strain
H3119F05	BG073178	Midline 1	Mid1	17318	X	XVC	4.84E-05	0.04	-3.41	7.87	6.10	J
H3110D11	BG072431						7.05E-05	0.04	-3.30	6.55	4.83	J
H3078C03							2.89E-05	0.03	-3.28	7.64	5.92	J
H3073A11	BG069130						4.33E-05	0.04	-2.90	5.59	4.05	J
AI853889	AI853889	Transcribed locus			15	VC	4.74E-07	0.00	-2.78	8.42	6.95	J
H3040B06	BG066202	Ubiquitin specific peptidase 24	Usp24	329908	4		4.51E-05	0.04	-2.73	5.57	4.12	J
H3017G07	BG064282	Placenta specific 9	Plac9	211623	14	VC	2.81E-07	0.00	-2.63	9.33	7.93	J
AI839337	AI839337					XHVC	8.93E-08	0.00	-2.30	8.86	7.66	J
H4025D12	BQ553995	Lamin B1	Lmnb1	16906	18		5.84E-05	0.04	-2.10	5.96	4.89	J
AI852313	AI852313	Gap junction membrane channel protein alpha 1	Gja1	14609	10		1.01E-05	0.02	-2.08	7.26	6.21	J
AI845815	AI845815	RIKEN cDNA 5430432M24 gene	5430432M24Rik	73847	2		6.46E-05	0.04	-1.96	7.34	6.37	J
AI844710	AI844710	Transcribed locus			14		1.08E-05	0.02	-1.87	7.57	6.67	J
AI851117	AI851117	Signal recognition particle 72	Srp72	66661	5		3.38E-05	0.03	-1.80	8.10	7.25	J
H3106G05	BG072101						6.46E-05	0.04	-1.79	6.67	5.83	J
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	67464	14	HVC	4.15E-06	0.01	-1.78	10.32	9.49	J
H3108G07	BG072279	DNA segment, Chr 14, ERATO Doi 449, expressed	D14Ertd449e	66039	14	XVC	8.52E-07	0.00	-1.76	9.83	9.02	J
AI834942	AI834942	Transcribed locus			X		3.71E-05	0.04	-1.76	7.40	6.59	J
H3032F05						HV	7.54E-05	0.04	-1.72	7.00	6.21	J
H3151C12		RIKEN cDNA 5830428H23 gene	5830428H23Rik	414758	19		1.02E-05	0.02	-1.69	6.35	5.59	J
AI849820	AI849820					VC	4.39E-06	0.01	-1.63	8.14	7.43	J
AI853572	AI853572	Transcribed locus, strongly similar to XP_916059.1 similar to Cyclin I [Mus musculus]			11	HV	1.40E-05	0.02	-1.63	8.05	7.34	J
AI850485	AI850485	Cysteine and histidine-rich domain (CHORD)-containing, zinc-binding protein 1	Chordc1	66917	9		3.18E-05	0.03	-1.51	8.35	7.75	J
H3063G06	BG068284	Transcribed locus			18		2.28E-05	0.03	-1.49	7.21	6.64	J
H3032E04	CK334462	Kinesin family member 5B	Kif5b	16573	18		4.79E-05	0.04	-1.44	11.45	10.93	J
H3082H09	BG082965	Cyclin-dependent kinase 8	Cdk8	264064	5		8.62E-05	0.05	-1.39	8.99	8.51	J
AI847941	AI847941	Progesterone receptor membrane component 2	Pgrmc2	70804	3		5.36E-05	0.04	-1.32	9.14	8.73	J
AI849350	AI849350	ATPase, Ca++ transporting, plasma membrane 1	Atp2b1	67972	10	VC	8.24E-05	0.05	-1.31	11.34	10.95	J
H4066G07	BQ560834	RIKEN cDNA 0610007P22 gene	0610007P22Rik	68327	17		2.38E-06	0.01	-1.29	10.02	9.65	J
H3025D01	BG064920	Brain expressed gene 1	Bex1	19716	X	VC	7.46E-05	0.04	-1.26	11.18	10.85	J
AI837333	AI837333	Transcribed locus			10		4.19E-06	0.01	-1.26	11.89	11.55	J
AI841702	AI841702	Hexamethylene bis-acetamide inducible 1	Hexim1	192231	11		3.59E-06	0.01	-1.24	9.73	9.42	J
H4037G10	BQ556057	Kinesin family member 5A	Kif5a	16572	10		1.08E-06	0.00	-1.24	12.55	12.23	J
H3024C09	BG064827	Splicing factor, arginine/serine-rich 3 (SRp20)	Sfrs3	20383	17		7.01E-05	0.04	-1.18	10.53	10.30	J
AI848744	AI848744	Heat shock protein 90kDa alpha (cytosolic), class B member 1	Hsp90ab1	15516	17		3.37E-05	0.03	1.10	13.02	13.16	NCrl
406125	AI413860	Kv channel interacting protein 3, calsenilin	Kcnip3	56461	2		2.57E-06	0.01	1.16	9.91	10.13	NCrl
AI847697	AI847697	HIG1 domain family, member 2A	Higd2a	67044	13		4.35E-05	0.04	1.17	11.74	11.96	NCrl
H3058D09	BG067809	Guanine nucleotide binding protein (G protein), gamma 3 subunit	Gng3	14704	19		2.32E-06	0.01	1.22	12.18	12.46	NCrl
479101	AI427473	COP9 (constitutive photomorphogenic) homolog, subunit 8 (Arabidopsis thaliana)	Cops8	108679	1		4.22E-05	0.04	1.24	9.15	9.47	NCrl
H3031D05	BG065459	Coatomer protein complex, subunit gamma	Copg	54161	6		4.62E-05	0.04	1.26	9.96	10.29	NCrl
H3006E01							1.69E-05	0.02	1.26	7.44	7.78	NCrl
AI844098	AI844098	Proprotein convertase subtilisin/kexin type 2	Pcsk2	18549	2		2.08E-05	0.02	1.34	10.23	10.66	NCrl
AI848715	AI848715	Adenosine A1 receptor	Adora1	11539	1		4.47E-05	0.04	1.35	8.84	9.28	NCrl
H4035C12	BQ555651	Potassium large conductance calcium-activated channel, subfamily M, beta member 4	Kcnmb4	58802	10		1.73E-05	0.02	1.48	8.18	8.75	NCrl
H3042G10	BG066421	Necdin	Ndn	17984	7		6.24E-05	0.04	1.48	11.06	11.63	NCrl
AI836976	AI836976	Glutamate dehydrogenase 1	Glud1	14661	14		4.06E-05	0.04	1.52	8.31	8.92	NCrl
H3090G11	BG070756	Zinc finger CCCH type containing 14	Zc3h14	75553	12		8.90E-05	0.05	1.54	7.00	7.62	NCrl
H3076F11	BG069378	Potassium voltage-gated channel, subfamily H (eag-related), member 3	Kcnh3	16512	15		5.65E-05	0.04	1.66	7.14	7.87	NCrl
AI844093	AI844093	Transcribed locus			12		1.59E-05	0.02	1.77	6.18	7.00	NCrl
H4033E01	BQ555328	WD repeat and FYVE domain containing 1	Wdfy1	69368	1		1.43E-06	0.00	2.73	6.50	7.94	NCrl
AI838156	AI838156	Carbonic anhydrase 8	Car8	12319	4		1.22E-06	0.00	3.11	7.93	9.57	NCrl
AI843793	AI843793	Purkinje cell protein 2 (L7)	Pcp2	18545	8		1.05E-06	0.00	4.70	7.36	9.59	NCrl

Table 3.4: Transcripts highly significantly divergent between substrains in the ventral brain region. A small number of genes are highly statistically significantly divergent between B6 substrains based on the criterion $Q<0.05$. (CLID, clone ID; GBID, GeneBank Acession; Entrez, Entrez ID; Chr, chromosome; X, cortex and striatum; H, Hippocampus; V, Ventral brain region; C, cerebellum.) Top panel is cortex and striatum and bottom panel is hippocampus.

Cerebellum

CLID	GBID	Name	Symbol	Entrez	Chr	Multiple Regions	Cer P	Cer Q	Cer FC	B6J AVG	B6NCrl AVG	Strain
H3017G07	BG064282	Placenta specific 9	Plac9	211623	14	VC	6.81E-09	0.00	-3.47	8.82	7.02	J
H3119F05	BG073178	Midline 1	Mid1	17318	X	XVC	3.62E-06	0.01	-3.05	7.90	6.29	J
AI851393	AI851393	Rhomboid domain containing 2	Rhbdd2	215160	5		1.93E-05	0.02	-2.71	6.93	5.49	J
H3035A06							5.60E-06	0.01	-2.68	8.76	7.33	J
AI839337	AI839337					XHVC	1.34E-09	0.00	-2.68	8.24	6.82	J
AI853889	AI853889	Transcribed locus			15	VC	4.24E-05	0.04	-2.45	7.65	6.36	J
H3108G07	BG072279	DNA segment, Chr 14, ERATO Doi 449, expressed	D14Ertd449e	66039	14	XVC	7.88E-07	0.00	-2.28	9.63	8.44	J
AI849820	AI849820					VC	1.89E-10	0.00	-2.28	9.18	7.99	J
H4012H09	BQ552019						4.61E-07	0.00	-2.25	9.85	8.68	J
AI852578	AI852578	Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	Slc11a2	18174	15		5.28E-05	0.04	-2.11	7.51	6.44	J
AI838331	AI838331	Transcribed locus			4		7.70E-06	0.01	-2.09	10.01	8.95	J
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	67464	14	HVC	1.18E-07	0.00	-2.07	10.21	9.16	J
H4006C12	BQ550938	Solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 6	Slc17a6	140919	7		4.67E-05	0.04	-2.01	7.32	6.32	J
H3080B09							1.81E-05	0.02	-1.85	6.16	5.28	J
H3027F01							4.06E-06	0.01	-1.58	8.35	7.69	J
H3086D07	BG070351	TatD DNase domain containing 1	Tatdn1	69694	15		2.49E-07	0.00	-1.49	6.37	5.79	J
AI847844	AI847844	Transcribed locus			13		2.13E-06	0.01	-1.40	8.13	7.64	J
H4066B04	BQ560728						1.70E-05	0.02	-1.38	9.91	9.45	J
AI849350	AI849350	ATPase, Ca++ transporting, plasma membrane 1	Atp2b1	67972	10	VC	1.57E-06	0.00	-1.37	11.18	10.73	J
H3025D01	BG064920	Brain expressed gene 1	Bex1	19716	X	VC	2.83E-05	0.03	-1.33	11.28	10.87	J
AI845982	AI845982	Asparaginyl-tRNA synthetase	Nars	70223	18		4.91E-05	0.04	-1.32	11.22	10.82	J
316656	AI413992	Brain expressed X-linked 2	Bex2	12069	X		2.11E-06	0.01	-1.27	12.94	12.59	J
H3026E01							1.36E-05	0.02	-1.26	12.30	11.97	J
H3029A05	BG065256	Chromosome segregation 1-like (S. cerevisiae)	Cse1l	110750	2		1.87E-05	0.02	-1.23	8.30	8.01	J
H3124F03	CK334935	Prothymosin alpha	Ptma	19231	1		2.69E-06	0.01	-1.19	11.75	11.50	J
H3120D10	BG073248	Ribosomal protein S4, X-linked	Rps4x	20102	X		3.43E-05	0.03	-1.19	12.75	12.50	J
H3144E09	BG075206	Ribosomal protein L6	Rpl6	19988	5		6.36E-05	0.05	-1.17	12.75	12.51	J
H3108A04	BG072209	Clusterin	Clu	12759	14		1.31E-06	0.00	-1.11	12.75	12.59	J
AI839859	AI839859						5.43E-05	0.04	1.21	10.87	11.14	NCrl
481481	AI426269	Calcium/calmodulin-dependent protein kinase II gamma	Camk2g	12325	14		1.91E-05	0.02	1.22	10.04	10.33	NCrl
444235	AI426077	MAP/microtubule affinity-regulating kinase 3	Mark3	17169	7		6.07E-05	0.05	1.22	9.75	10.03	NCrl
AI837106	AI837106	Enolase 2, gamma neuronal	Eno2	13807	6		6.58E-05	0.05	1.27	12.58	12.93	NCrl
H4074C05	BQ562125	Serine (or cysteine) peptidase inhibitor, clade I, member 1	Serpini1	20713	3		7.18E-06	0.01	1.28	9.58	9.94	NCrl
H3151F10	BG075782	Protein phosphatase 1, catalytic subunit, gamma isoform	Ppp1cc	19047	5		1.38E-05	0.02	1.32	12.00	12.40	NCrl
H3023E06	BG064757	Drebrin-like	Dbnl	13169	11		3.65E-05	0.03	1.34	8.34	8.76	NCrl
337805	AI415313	Calcium channel, voltage-dependent, beta 3 subunit	Cacnb3	12297	15		4.26E-05	0.04	1.39	9.22	9.70	NCrl
AI843192	AI843192	Potassium voltage-gated channel, shaker-related subfamily, beta member 3	Kcnab3	16499	11		5.89E-05	0.05	1.51	8.10	8.69	NCrl
573351	AI323584	Glutaryl-Coenzyme A dehydrogenase	Gcdh	270076	8		1.61E-05	0.02	1.59	6.45	7.13	NCrl
H3125H04	BG073770	MAD2L1 binding protein	Mad2l1bp	66591	17		1.61E-05	0.02	1.63	5.53	6.23	NCrl
H3136F09	BG087195						1.28E-06	0.00	1.81	4.97	5.83	NCrl
596225	AI447280	Gremlin 1	Grem1	23892	2		3.46E-05	0.03	1.91	6.04	6.97	NCrl
H3092D03	BG070807	Transcribed locus			5		1.15E-05	0.02	1.97	4.91	5.89	NCrl
H3137D10	BG074632	Transmembrane protein 16F	Tmem16f	105722	15		6.66E-05	0.05	2.57	4.27	5.63	NCrl

Table 3.5: Transcripts highly significantly divergent between substrains in the cerebellum. A small number of genes are highly statistically significantly divergent between B6 substrains based on the criterion $Q<0.05$. (CLID, clone ID; GBID, GeneBank Acession; Entrez, Entrez ID; Chr, chromosome; X, cortex and striatum; H, Hippocampus; V, Ventral brain region; C, cerebellum.) Top panel is cortex and striatum and bottom panel is hippocampus.

Ventral Brain		
Biological Process	**p-Value**	**Gene symbol**
Transmembrane receptor protein kinase signaling pathway	4.28E-03	*Glud1*, **Bex1**, *Ndn*
Nerve growth factor receptor signaling pathway	1.27E-04	**Bex1**, *Ndn*
Microtubule-based process	8.28E-03	**Kif5b**, **Kif5a**, **Mid1**
Metal ion transport	3.62E-03	*Kcnb4*, *Kcnh3*, **Atp2b1**, *Kcnip3*
Potassium ion transport	2.09E-03	*Kcnb4*, *Kcnh3*, *Kcnip3*
Circulation	6.20E-03	*Kcnmb4*, **Gja1**
Sensory perception of pain	5.52E-04	*Ndn*, *Kcnip3*
Neuron migration	4.77E-03	*Ndn*, **Gja1**
Molecular Function	**p-Value**	**Gene symbol**
Specific transcriptional repressor activity	3.93E-04	**Hexim1**, *Kcnip3*
Channel or pore class transporter activity	5.36E-03	*Kcnmb4*, *Kcnh3*, **Gja1**, *Kcnip3*
alpha-type channel activity	3.99E-03	*Kcnmb4*, *Kcnh3*, **Gja1**, *Kcnip3*
Cation channel activity	8.42E-03	*Kcnmb4*, *Kcnh3*, *Kcnip3*
Potassium channel activity	1.71E-03	*Kcnmb4*, *Kcnh3*, *Kcnip3*
Cellular Component	**p-Value**	**Gene symbol**
Cell projection part	5.26E-03	**Kif5b**, **Kif5a**
Cilium	1.50E-03	**Kif5b**, **Kif5a**
Cilium part	3.56E-04	**Kif5b**, **Kif5a**
Ciliary rootlet	1.92E-04	**Kif5b**, **Kif5a**
Cytoskeletal part	6.99E-03	**Kif5b**, **Kif5a**, *Ndn*, **Lmnb1**, **Mid1**
Microtubule cytoskeleton	8.46E-03	**Kif5b**, **Kif5a**, *Ndn*, **Mid1**
Kinesin complex	4.28E-03	**Kif5b**, **Kif5a**

Cerebellum		
Biological Process	**p-Value**	**Gene symbol**
Regulation of cell organization and biogenesis	7.48E-03	**Mid1**, *Dbnl*
Negative regulation of cell organization and biogenesis	2.78E-03	**Mid1**, *Dbnl*
Negative regulation of protein metabolism	8.62E-03	**Mid1**, *Dbnl*
Ion transport	2.83E-03	**Atp2b1**, **Slc11a2**, *Kcnab3*, *Camk2g*, *Cacnb3*
Cation transport	6.28E-04	**Atp2b1**, **Slc11a2**, *Kcnab3*, *Camk2g*, *Cacnb3*
di-, tri-valent inorganic cation transport	6.72E-05	**Atp2b1**, **Slc11a2**, *Camk2g*, *Cacnb3*
Metal ion transport	1.54E-04	**Atp2b1**, **Slc11a2**, *Kcnab3*, *Camk2g*, *Cacnb3*
Calcium ion transport	3.34E-04	**Atp2b1**, *Camk2g*, *Cacnb3*
Protein depolymerization	2.61E-03	**Mid1**, *Dbnl*
Molecular Function	**p-Value**	**Gene symbol**
Cation transporter activity	6.34E-03	**Atp2b1**, **Slc11a2**, *Kcnab3*, *Cacnb3*
di, tri-valent inorganic cation transporter activity	1.86E-03	**Atp2b1**, **Slc11a2**
Metal ion transporter activity	3.12E-03	**Atp2b1**, **Slc11a2**
Cellular Component	**p-Value**	**Gene symbol**
Cytosolic part	1.69E-03	*Eno2*, **Rpl6**, **Rps4x**
Cytosolic ribosome	8.13E-03	**Rpl6**, **Rps4x**

Table 3.6: Overrepresented functional groups between substrains in the midbrain and cerebellum. Significant overrepresentation determined using the GOTree function in WebGestalt. Bold lettering and italics indicate genes with higher expression and lower expression respectively in B6J compared to B6C.

DISCUSSION

The purpose of this study is to both recommend caution when using inbred substrains and to offer other researchers tools in which to further study the genetic differences between B6J and B6C. We provide evidence in the form of brain region transcriptome analysis and by a phenotypic survey of alcohol related measures that B6J and B6C are non-identical at the genetic level. For alcohol researchers, it is important to be cognizant of the differences in voluntary alcohol intake between B6J and B6C. B6C displays a reduced preference for alcohol and a corresponding reduction in alcohol intake compared to the B6J substrain. It appears that this difference is not due to differences in alcohol metabolism, innate taste discrepancies or whether the animal was purchased from a vendor or bred in house. Researchers attempting to substitute B6J and B6C should also be aware of the many phenotypic and genotypic differences that have been documented previously as well as in this study.

Our brain region microarray analysis of vendor purchased B6J and B6C mice was deliberate so as to provide a valuable dataset for all researchers who might now wish to use the two substrains as a genetic model to test novel parameters for any of the specific candidate genes we have generated. In fact, numerous significantly divergent transcripts of diverse biological function were identified, confirming again that B6J and B6C differ at the level of the genome in many ways. It is also important to note that although we did detect many significant changes in gene expression, we did not find a specific transcription factor, transcription factor binding site or enhancer/repressor region that could account for the bulk of the gene expression differences observed between the substrains. We suggest that the primary reason for the divergence is due, in fact, to polymorphisms in the genome. However, direct confirmation is needed.

In this study, several chromosome regions of interest were identified for further investigation. Four chromosomal regions that contained clusters of significantly co-regulated and co-localized genes were observed. The localization of these genes to clusters is non-random and indicates a higher level of chromosome regulation operating through an undetermined mechanism, possibly through differences in chromatin structure. We were also able to identify regions on chromosome 14 and chromosome 4 that appear to contain copy number gains in the B6J genome that are not represented in the B6C substrain. An analysis of inbred strain expression data and comparative genomic scans for the chromosome 14 region indicated that it is likely that B6C mice may be lacking 1-2 copies of the genes within the region. A similar comparison within the chromosome 4 region, including Ccl21b and 4933409K07Rik, found that B6C mice may be missing 1 copy of the representative genes from that region although further validation will be required. Loss of extra copies of genes is not unusual and can easily occur through mismatched crossing over during meiosis, in fact, it appears that similar events may have occurred in other inbred strains. Although the representative genes from the chromosome 14 region (*D14Ertd449e* and *Plac9*) are abundantly expressed in brain, their function remains unknown at this time and warrants further investigation.

Based on the different levels of alcohol consumption and alcohol preference observed between B6J and B6C we attempted to find genes within our most significant microarray that had been previously associated with alcohol consumption or a response to alcohol. Using this approach we identified 31 genes (table 3.2) that were found to differ between both B6J and B6C and high alcohol and low alcohol consuming inbred strains of mice (Mulligan, Ponomarev et al. 2006). Nearly 70% of these genes are expressed higher in B6J compared to B6C animals which indicates that B6J, compared to B6C, contains more genes associated with higher alcohol consumption and increased alcohol preference

(Mulligan, Ponomarev et al. 2006). Genes previously identified from other studies as being associated with alcohol related behaviors or addiction included the serotonin transporter, Slc6a4, nerve growth factor, Gap43, a potassium channel beta subunit, Kcnmb4, the vesicular glutamate transporter, Slc17a6, and Pcsk2, which is required for Cocaine- and amphetamine-regulated transcript (CART) peptide processing. All of these genes are candidates for regulating the phenotype of reduced alcohol intake observed in B6C relative to B6J.

Alcohol abuse and dependence is a pervasive problem in human populations around the globe. Despite decades of research many of the genes that underlie these maladaptive behaviors remain unknown. By making brain transcriptome information available for the B6J and B6C substrains we hope to provide researchers with the tools necessary to address the genetic components of high alcohol consumption. Serendipitously, these two sister substrains also represent a novel model system in which to study the divergence of inbred populations over time and the resulting effect these genetic changes have on phenotype.

Table A Genes that Overlap with Alcohol Preferring and Alcohol Non-Preferring Inbred Strains of Mice

CLID	GBID	Name	Symbol	Strain	Region	Evidence
H4037B11	BQ555948	Nudix (nucleotide diphosphate linked moiety X)-type motif 3	Nudt3	B6J	C&S	Mulligan, Ponomarev et al., 2006
AI851117	AI851117	Signal recognition particle 72	Srp72	B6J	Vb	Mulligan, Ponomarev et al., 2006
AI847941	AI847941	Progesterone receptor membrane component 2	Pgrmc2	B6J	Vb	Mulligan, Ponomarev et al., 2006
H3108A04	BG072209	Clusterin	Clu	B6J	Cer	Mulligan, Ponomarev et al., 2006; Murphy et al., 2002
AI852578	AI852578	Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	Slc11a2	B6J	Cer	Mulligan, Ponomarev et al., 2006
H3124F03	CK334935	Prothymosin alpha	Ptma	B6J	Cer	Mulligan, Ponomarev et al., 2006
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	B6J	All	Mulligan, Ponomarev et al., 2006
H3045G06	BG066593	Translocated promoter region	Tpr	B6J	All	Mulligan, Ponomarev et al., 2006
H3066H07	BG081591	Heme oxygenase (decycling) 2	Hmox2	B6J	All	Mulligan, Ponomarev et al., 2006
AI854596	AI854596	Integrin linked kinase	Ilk	B6J	All	Mulligan, Ponomarev et al., 2006
H4064F12	BQ560502	Presenilin 1	Psen1	B6J	All	Mulligan, Ponomarev et al., 2006
AI846695	AI846695	Quaking	Qk	B6J	All	Mulligan, Ponomarev et al., 2006
H4066C12	BQ560762	Mitogen-activated protein kinase kinase kinase kinase 4	Map4k4	B6J	All	Mulligan, Ponomarev et al., 2006
AI835853	AI835853	H2A histone family, member Z	H2afz	B6J	All	Mulligan, Ponomarev et al., 2006
AI846370	AI846370	Methionine aminopeptidase 2	Metap2	B6J	All	Mulligan, Ponomarev et al., 2006
H4064E12	BQ560480	Sphingomyelin phosphodiesterase 3, neutral	Smpd3	B6J	All	Mulligan, Ponomarev et al., 2006
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	B6J	All	Mulligan, Ponomarev et al., 2006
H4004A11	BQ550561	N-6 adenine-specific DNA methyltransferase 2 (putative)	N6amt2	B6J	All	Mulligan, Ponomarev et al., 2006
H4064D04	BQ560445	RIKEN cDNA 2810405J04 gene	2810405J04Rik	B6J	All	Mulligan, Ponomarev et al., 2006
H4010H01	BQ551677	Debranching enzyme homolog 1 (S. cerevisiae)	Dbr1	B6J	All	Mulligan, Ponomarev et al., 2006
H3072D08	BG069071	Ectonucleoside triphosphate diphosphohydrolase 4	Entpd4	B6J	All	Mulligan, Ponomarev et al., 2006
423422	AI425884	Tripartite motif protein 10	Trim10	B6NCr1	All	Mulligan, Ponomarev et al., 2006
H4003D06	BQ550441	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1	Yes1	B6NCr1	All	Mulligan, Ponomarev et al., 2006
475847	AI427464	RIKEN cDNA 2410127L17 gene	2410127L17Rik	B6NCr1	All	Mulligan, Ponomarev et al., 2006
H4033E01	BQ555328	WD repeat and FYVE domain containing 1	Wdfy1	B6NCr1	All	Mulligan, Ponomarev et al., 2006

Table B Genes Previously Found to be Associated with Alcohol or Addiction Phenotypes

CLID	GBID	Name	Symbol	Strain	Region	Evidence
H3157F06	BG076269	Growth associated protein 43	Gap43	B6J	C&S	Perrone-Bizzozero, Isaacson et al. 1998 and Kim, Choi et al. 2006
AI846650	AI846650	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	Slc6a4	B6J	H	Boyce-Rustay et al., 2006 and Lesch 2005
H4006C12	BQ550938	Solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 6	Slc17a6	B6J	Cer	Zhou et al., 2006
H4035C12	BQ555651	Potassium large conductance calcium-activated channel, subfamily M, beta member 4	Kcnmb4	B6NCr1	Vb	Feinberg-Zadek and Treistman 2007
AI844098	AI844098	Proprotein convertase subtilisin/kexin type 2	Pcsk2	B6NCr1	Vb	Salinas, Wilde et al. 2006; Dayas, McGranahan et al. 2007 and Dandekar, Singru et al. 2007
H4067F11	BQ560996	Solute carrier family 4, sodium bicarbonate cotransporter, member 7	Slc4a7	B6NCr1	All	Ishiguro, Walther et al. 2007

Table 3.7: Genes previously known to be associated with alcohol consumption or alcohol or addiction phenotypes. (A) Overlap between high expression in high alcohol consuming strains (versus low consuming strains) and high expression in B6J compared to B6C or overlap between low alcohol consuming strains (versus high consuming strains) and high expression in B6C compared to B6J. Only a few genes overlap with the later comparison. Evidence is the citation for the publication on which the overlap analysis is based. Clusterin was also previously identified as being ethanol responsive. (B) Genes associated with alcohol or addiction phenotypes. Evidence is the citation for the publication for which a link to an alcohol or addiction phenotype was observed.

MATERIALS AND METHODS

Animal handling at the Jackson Laboratories

C57BL/6J mice colonies from the Jackson Bar Harbor facility were fed *ad libitum* water and NIH Mouse/Auto 6F 5K52 diet consisting of not less than 18% crude protein, 6% crude fat, 5% crude fiber, 8% ash and 3% added minerals. The light cycle was 12:12 and all other environmental controls were standard. The breeding protocol at the Jackson Laboratory used a trio of 2 females and 1 male, with offspring weaned at 28 days. All mice purchased directly from Jackson Labs were shipped at approximately 56 days and held for 2 weeks or more prior to testing at the University of Texas at Austin.

Animal handling at the Charles River Laboratories

Breeding also occurred using a breeding trio, and offspring are weaned at 21 days. Mice were fed a diet of Purina 5L79 which consists of 18% crude protein, 5.2% crude fat, 5.2% crude fiber, 5.7% Ash and 3.3% minerals (Cano, Arnold et al. 2001). All mice from Charles River Labs were shipped at approximately 56 days and held for approximately 21 days prior to testing at the University of Texas at Austin on a 12 hour light/dark cycle.

Animal handling at the University of Texas at Austin

B6 mice were bred in house from adult male and female mice (1:1) supplied from both Jackson (B6J) and Charles River Laboratories (B6C). The offspring will be designated B6JUT and B6CUT in order to differentiate them from the vendor purchased animals. B6JUT and B6CUT mice were weaned at 21 days and fed a diet of PROLAB RMH 1800 which contains a minimum of 18% crude protein, 5% minimum crude fat and 5% minimum crude fiber. Weaned animals were housed with up to 5 same sex animals

on a 12 hour light/dark cycle until experimentation began after they were at least 9 weeks of age.

Alcohol Consumption and Preference Measurement

Vendor purchased mice

Ten individually housed adult female mice from each vendor, were provided with a sipper tube containing an ethanol and water solution and a sipper tube containing tap water. In this two-bottle choice paradigm animals were allowed free access to food, alcohol solutions and water. Starting at a 3% alcohol and water solution (w/v) for 4 days and increasing in increments of 3% up to 21% alcohol and water (w/v), animals were allowed access to each additional concentration for 5 days before being switched to the next highest concentration. The physical location (right or left) of the water and alcohol solutions was switched daily to reduce the effects of potential side preference. Sipper tubes were weighed daily to the nearest 0.01 grams in order to measure consumption of alcohol and water. To avoid including aberrant data from leaky tubes or from mice that played with their bottles, an outlier test was conducted on the daily total fluid intake and outliers were excluded on a per day basis. Consumption was determined as grams of alcohol consumed per kilogram body weight per day. Preference was measured as the amount of alcohol consumed over the total amount of fluid consumed per day, with >50% designated as a preference for alcohol. The statistics software GraphPad Prizm (Jandel Scientific, Costa Madre, CA) was used to complete ANOVAs to evaluate the effect of strain on consumption of alcohol and alcohol preference.

Mice bred in house

Adult mice from each substrain and both sexes were tested using the same two-bottle choice paradigm described above with the exception that mice were tested for 10

days using a 20% alcohol and water solution (n=12 females per group, and n=6 and 7 respectively for JUT and CUT male mice). Differences in consumption and preference between the strains were tested for significance by a *t*-test within each sex.

Taste discrimination

Vendor purchased mice

Taste preference was measured using a saccharine and quinine two-bottle choice paradigm similar to that described above. Female adult mice from each strain were housed individually in single cages and provided with two sipper tubes containing either a 0.033% saccharine solution and tap water or a 0.03mM quinine solution and tap water. Consumption was measured for 4 days and bottles were alternated each day to avoid side preference effects.

Loss of righting reflex

Vendor purchased mice

Ten B6J and 10 B6C adult female animals were injected with 3.8g/kg ethanol intraperitoneally (i.p.). Ethanol-induced loss of righting reflex was measured by placing mice on their backs in a ~90° angle plastic trough. Loss of righting reflex (LORR) was described as the inability of the mouse to right itself within 30 sec. Return of the righting reflex was operationally defined as the ability of the mouse to right itself twice in a 1 minute period. The duration of LORR was measured as the time between the loss of the righting reflex and the return of the righting reflex. Upon recovery of the righting reflex, retroorbital sinus blood was taken for gas chromatographic determination of blood alcohol concentration (BAC). The statistical software GraphPad Prizm was used to complete t-tests to evaluate the effect of strain on LORR and BAC.

Initial sensitivity to alcohol

Vendor purchased mice

Nine adult female mice of each substrain were used. Initial sensitivity to the hypnotic effects of ethanol was assessed using a modified version of LORR (Ponomarev and Crabbe 2002; Ponomarev and Crabbe 2004). Briefly, an animal was injected with 3 g/kg ethanol i.p. (20% v/v in saline) and then immediately placed in a cylindrical restrainer. The restrainer was then gently rotated 90 degrees every 2-3 seconds. Within 2 minutes of the injection the mice are heavily intoxicated and remain on their back after two successive 90 degree turns. Thus, the mouse is considered to have lost its righting reflex when it is no longer able to right itself within 5 seconds from a supine position. Latency to LORR in seconds was used as a measure of initial sensitivity.

Custom cDNA microarrays

cDNA microarrays were printed in house at the University of Texas, exactly as previously published (Mulligan, Ponomarev et al. 2006). The spotted clones on the microarrays represented approximately 13,000 unique known genes (annotations current as of June 29, 2007 using SOURCE at http://source.stanford.edu/cgi-bin/source/sourceSearch). Specific details for custom spotted cDNA microarray printing has been described elsewhere (Schena, Shalon et al. 1995).

Brain tissue collection and RNA isolation

Naïve adult animals from each supplier were sacrificed by cervical dislocation and the cortex (including striatum), hippocampus, cerebellum and ventral brain region (remaining brain region) were dissected. The olfactory bulb was excluded from this study. Dissected tissue was immediately frozen in liquid nitrogen and stored at –80 C. RNA was extracted according to the manufacturer's protocol using RNA STAT-60™

(Tel-Test, Inc., Friendswood, TX). RNA concentration and integrity were determined using the NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Agilent 2100 Bioanylizer and RNA 6000 Nano LabChip kit (Agilent Technologies, Palo Alto, CA) respectively.

Microarray hybridization and normalization

Total RNA for each brain region was used in subsequent microarray experiments. Microarray hybridizations were performed according to the manufacturer's protocol using the Genisphere (Hatfield, PA) Array 350 kit using 3 ug per sample. All experimental samples were end-labeled with Cy-3 (green channel 1) and hybridized against a common reference sample (total RNA from 100 adult male C57BL/6J mice) that was end-labeled with Cy-5 (red channel 2). Use of a common reference resulted in a total of 64 microarrays completed; 8 arrays for each B6 substrain in each of the 4 brain regions. Hybridized microarrays were immediately scanned using Axon GenePix 4000B dual channel laser scanners (Molecular Devices, Union City, CA). Microarrays were gridded using the software package GenePix version 5.0 (Molecular Devices). Lowess within-print-tip normalization was performed for each array using the R package, Statistics for Microarray Analysis (Yang, Dudoit et al. 2002). All normalized intensity information from the arrays used in this study were uploaded and stored in the Longhorn Array Database (Killion, Sherlock et al. 2003).

Microarray analysis

Red channel 2 (Ch2) and green channel 1 (Ch1), net median intensities were downloaded from LAD and the Ch1 intensities for each gene were adjusted (standardized) by the reference Ch2 intensities for each gene as follows. All intensity values less than 0 were given a value of 1. All linear intensity information was then

converted to $\log_2$ intensity values. For every gene on the microarray and for each experiment: $Ch1_S = Ch1 + (Ch2_{med} - Ch2)$, where $Ch1_S$ represents the standardized Ch1 intensity for an experimental value of a gene, Ch1 and Ch2 are the original green and red intensities, respectively, that represent the experimental value for that gene and $Ch2_{med}$ is the median value of the red channel intensities for that gene across all experiments. Outliers were removed by gene within each experimental group from the resulting normalized and standardized Ch1 intensity data. Two filters were applied to the microarray dataset. First, a filter was applied that removed genes that did not meet a minimum criteria of 6 (out of 8) Ch1 intensity values for each experimental group. Second, genes that did not exceed a minimum average Ch1 intensity threshold of 4.64 ($\log_2$ scale) across all experimental groups tested were discarded on the basis that the signal was below the threshold for biological relevance to be determined. Statistical significance was determined for each gene between experimental groups by *t*-test and false discovery rate estimation in the form of Storey's Q-value (Storey and Tibshirani 2003) or additionally as described below. Both *p*-value and *q*-value (*q*) were calculated using the free statistical software program R (Ihaka and Gentleman, 1996). Highly significant transcripts were selected for each brain region based on $q < 0.05$.

Detection of similarly expressed genes in all four brain regions

Transcripts that passed the intensity and missing data filter were evaluated for consistency of divergent expression between B6J and B6C across all four brain regions. The Student's two-tailed *t*-test was used to compare B6J and B6C in each brain region. The resulting *t*-value (*t*) was used to estimate the effect size for each brain region as follows: Cohen's *d*: $d = 2t / \sqrt{df}$, where *d* is the effect size measure Cohen's *d* and *df* is the degrees of freedom. A two-tailed *Z*-test (*z*) was performed on the effect sizes to assess significance. In order for a transcript to be considered to be highly significantly

divergently expressed between B6J and B6C in all four brain regions, we defined the following criteria: $z<0.01$, average $|d| \geq 0.8$ and average fold change $\geq |1.5|$). Significant genes were clustered and dispayed using Cluster 3.0 (de Hoon, Imoto et al. 2004) and JavaTreeview (Saldanha 2004) software.

Validation of expression data

Quantitative real-time PCR

Probes for *D14Ertd449e* and *Plac9* were purchased from Applied Biosystems (Foster City, CA). One µg of total RNA was taken from each of the 4 brain regions from 4 B6J and 4 B6C mice and pooled together to create n= 4 whole brain RNA pools for each substrain. Validation was performed as described in (Ponomarev, Maiya et al. 2006) with the exception that 18s rRNA was used as the endogenous control.

Overrepresentation analyses

Overrepresentation for pathways, transcription factor binding sites, and chromosomal location were completed using WebGestalt, oPOSSUM, and a macro within Microsoft Excel® respectively. Genes with similar patterns of expression in each substrain for each brain region were analyzed for overrepresented pathways using GOTree in WebGestalt (Zhang, Schmoyer et al. 2004; Zhang, Kirov et al. 2005) and for transcription factor binding site (TFBS) overrepresentation using oPPOSUM (Ho Sui, Mortimer et al. 2005). A $p<0.01$ significance cut-off was applied in both cases.

To investigate the effects of genotype on chromosome location-specific gene expression we compared chromosomal frequencies of genes differentially expressed between B6J and B6C to a random distribution (transcripts included were regulated in the same direction across all four brain regions based on z-test $p<0.05$ and average $|d| \geq 0.5$). First, all transcripts that passed the expression detection threshold were sorted by their

gene symbols and the level of statistical significance. Then, transcripts with unknown gene symbols and all duplicated gene copies with lower statistical significance were removed. The remaining genes with unique genetic identifiers were then resorted according to their chromosomal location. The numbers of closely located genes significantly different between the two strains and regulated in the same direction were then calculated within 15-gene overlapping bins along the whole genome. To calculate p-values associated with these frequencies the frequencies were then compared to a distribution of about 1.2 million frequencies calculated the same way and based on the same number of significantly regulated genes, but distributed randomly along the whole genome. This distribution was generated using 100 genome-wise permutations and a random number generator in Microsoft Excel®. Based on a binomial distribution, the frequency of 7 closely located similarly regulated genes (out of 15 genes in a bin) was considered to pass the significance threshold with $p < 0.0002$ and FDR < 13%. FDR was calculated as a proportion of the number of values equal or greater than 7 predicted from the random distribution to the number of similar values observed in our dataset.

Chapter 4: Overdominant Inheritance Patterns of Expression in a FVB.B6 F1 Hybrid with High Levels of Alcohol Consumption

INTRODUCTION

Human addictive behaviors are complex quantitative traits that can be explained by the interaction of genes, the environment and the interaction between genes and the environment. For this reason, locating candidate genes for human behaviors such as drug and alcohol seeking can be quite difficult. Although alcoholism is a disease with a significant genetic component, many of the genes and gene interactions involved in the addiction process remain unknown. Rodent models have provided a basic approach to understanding the genetic underpinnings of components of the addiction pathway. The existence of many different inbred strains of mice provides an excellent genetic model system for testing the relationship between genes and alcohol related behaviors. Although human behaviors, like drug and alcohol addiction are too complex to model outright in a mouse, different components of the disease can be studied, such as the rewarding effects of ethanol or sensitivity or tolerance to the effects of alcohol. The use of inbred strains also offers an advantage over human populations for studying complex behavioral traits because inbred animals are homozygous at all alleles, differ in a range of phenotypes and, in an ideal laboratory setting, the environment can be controlled and manipulated.

In this study, as in the previous studies, the genetic predisposition for high alcohol consumption is being evaluated. The component of the addiction process involved in this model system is sensitivity to the reinforcing effects of alcohol. Many different techniques have been developed in which to measure different dimensions of the reinforcing effects of alcohol, including voluntary ethanol self-administration and conditioned place preference. Twenty-four hour, two-bottle choice ethanol preference

has been used in many laboratories since it was first used in the late 1950's to measure the preference for alcohol solutions of inbred strains of mice (McClearn and Rodgers 1959; Fuller 1964). In the two-bottle choice paradigm, preference for alcohol over other solutions can be thought of as a measure of the rewarding or reinforcing properties of alcohol. This technique has proven to be a simple yet reliable measure both across laboratories and over time for the assessment of alcohol reinforcement (Wahlsten, Bachmanov et al. 2006). Recent surveys of inbred strain alcohol preference using the two-bottle choice paradigm have shown that the strain with the highest preference for alcohol is C57BL/6J (B6) (Belknap, Crabbe et al. 1993). As a consequence of this discovery, C57BL/6 animals are widely used to study alcohol-related traits because they will consume large quantities of alcohol and show nearly complete preference for 10% solutions of alcohol compared to water (McClearn and Rodgers 1959). A recent study demonstrated that a B6 hybrid mouse (F1) that was the result of a cross between the high alcohol preferring B6 strain and a low alcohol preferring strain (FVB/NJ) showed even higher alcohol consumption than either parental strain (figure 4.1) (Blednov, Metten et al. 2005). The difference in alcohol consumption observed in the preceding study was likely not due to the effects of ethanol metabolism, taste preference or body weight. The F1 hybrid animals also showed alcohol preference and consumption similar to that of the B6 parental strain in another paradigm (drinking in the dark, or DID) designed to lead to high blood alcohol concentrations (BAC) in a short time period, suggesting that both strains share the genetic propensity for high voluntary alcohol consumption. The elucidation of the increase in two-bottle choice alcohol consumption in the F1 hybrid compared to the parental strains suggested the possibility that the behavior of high alcohol intake in these animals is the result of overdominance or epistasis and provided a framework for which to identify candidate genes underlying sensitivity to the reinforcing effects of alcohol.

The advancement of molecular and quantitative genetic techniques has lead to new attempts to investigate the interaction between genotype and phenotype. Microarray technology allows for the global measurement of gene expression in an infinite number of samples and will continue to play an integral role in understanding how genes and the environment interact in order to create a response. Understanding this interaction is the key to understanding disease development and maladaptive behavioral states, such as drug and alcohol addiction. Several recent studies have used trancriptome profiling approaches to identify patterns of inheritance in F1 model systems. The goal of such studies is to better understand the inheritance and regulation of transcript levels as well as the architecture of gene regulation. In terms of modes of inheritance, many studies using maize have attempted to explain the phenomenon of heterosis or hybrid vigor. For maize and other agricultural species, heterosis is the intriguing case in which the F1 hybrid of two inbred lines demonstrates increased performance compared to either parent. Current models for heterosis involve complementation of deleterious mutations by dominant genes or overdominance resulting from combinations of alleles that offer enhanced performance (Lippman and Zamir 2007). Although the phenotype of high alcohol drinking in the B6 hybrid, is not exactly the same phenotype as enhanced agronomic performance observed in maize F1 crosses, we can hypothesize that the same molecular events play a role in the enhancement of two-bottle choice alcohol consumption.

For any F1 hybrid system, two models of gene expression have been proposed in regards to inheritance patterns in the F1 (Gibson and Weir 2005). The first model assumes that hybrids will inherit the phenotype of gene expression in an additive manner. Additivity describes a transcript that is expressed close to the average of the two parental strains; this value is often described as the midparent level. The second model assumes that transcript expression will be non-additive or dominant. There are several different

modes of dominant expression; transcript expression in the F1 could be similar to one of the parental strains or F1 expression could be overdominant or underdominant to both parental strains in which case expression exceeds that of either parent or is less than each parental strain. The two models of inheritance are not mutually exclusive and both additive and dominant inheritance patterns have been observed in maize, mice and drosophila (Gibson, Riley-Berger et al. 2004; Cui, Affourtit et al. 2006; Stupar and Springer 2006; Swanson-Wagner, Jia et al. 2006). In some studies additivity exceeded all other forms of inheritance (Cui, Affourtit et al. 2006; Stupar and Springer 2006; Swanson-Wagner, Jia et al. 2006) and in other cases, inheritance was found to be primarily non-additive (Gibson, Riley-Berger et al. 2004). In order to investigate the genetic contributions to high alcohol consumption in the B6 F1 model, whole brain gene expression in the FVB, B6 and FVB.B6 hybrid was assayed and then tested for additive and non-additive patterns of expression in the FVB.B6 hybrid. Special attention was given to the genes that show under- and overdominant expression as they may play a role in the development of enhanced consumption compared to B6 in the two-bottle choice alcohol preference paradigm.

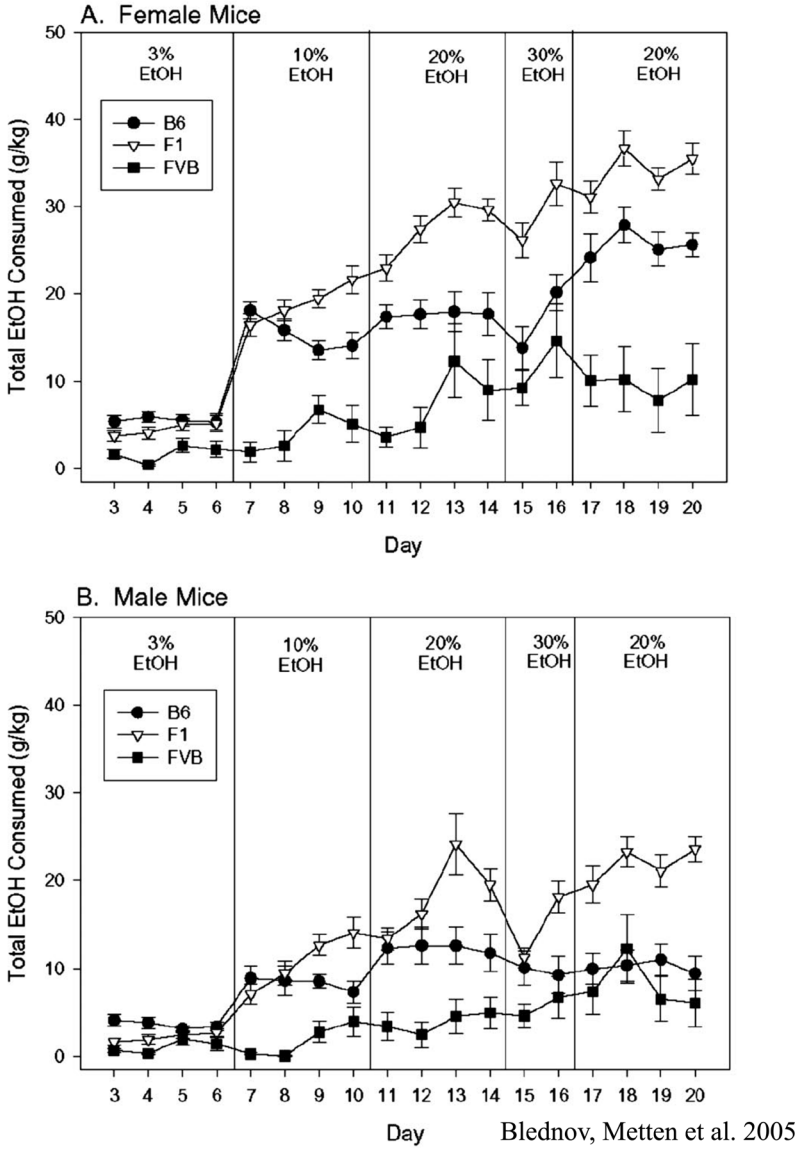


Figure 4.1: Consumption of alcohol solutions in C57BL/6, FVB/N and FVB.B6 hybrid strains. Consumption is measured in grams (g) of alcohol consumed per kilogram (kg) body weight. The x-axis shows the percentage of alcohol solution over time in days. Female (A) and male (B) FVB.F1 hybrid mice consume more alcohol than either parental strain. Figure published previously in Blednov, Metten at al. 2005.

RESULTS

Clustering revealed patterns of both additive and non-additive inheritance in the F1 hybrid

A total of 32,979 transcripts met the criteria for reliable expression. Of those transcripts, 3,012 were identified as significantly divergent in one or more of the three genotypes ($P<0.05$) by one-way ANOVA. Heirarchical clustering of the significant genes by strain (experiment type) revealed that the hybrid strain is more similar to the FVB parental strain than it is to the B6 parental strain (Figure 4.2). It is possible that maternal strain influences transcript inheritance in the F1 since FVB is the maternal strain in the F1 hybrid, however, without an analysis of brain gene expression in the reciprocal hybrid (B6.FVB) these results are not conclusive. It should be mentioned that there is no difference in two-bottle choice alcohol consumption between reciprocal B6.FVB and FVB.B6 hybrids (Blednov, Metten et al. 2005). Other studies in drosophila and mice found that global gene expression in the F1 reciprocal crosses was more similar to each other than to either parental strain (Gibson, Riley-Berger et al. 2004; Cui, Affourtit et al. 2006).

K-means clustering organizes gene expression data based on assignment of genes into a user defined number of clusters (k) with the goal of reducing the variance within each cluster. Figure 4.2 shows the results of K-means clustering of the 3,012 significantly divergent transcripts with k=6. Nearly a quarter of the transcripts were localized to two clusters that displayed overdominance and underdominance relative to parental strain expression (figure 4.3A). The remaining 4 clusters showed divergent expression patterns between the parental strains that were either additive or dominant in the F1 (figure 4.3B) The first two clusters in panel B show transcripts whose expression patterns are similar between the FVB parental strain and the F1 hybrid strain. The last

two clusters show a pattern of expression in the F1 that was more similar to the B6 strain. Although genes with additive patterns of expression in the F1 did not assemble into independent clusters from genes with dominant patterns of expression, (multiple values of k were attempted in the K-means algorithm) they can be visualized in all four clusters in panel B as black regions or regions with intermittent levels of red and green within each cluster (see Figure 4.3B). The clustering together of transcripts that have both additive and non-additive expression in the F1 revealed the broad range of expression patterns in the F1. The strongest factor in grouping expression patterns in this data set seemed to arise from the divergence in expression between the two parental strains. When there are large differences in expression between the parental strains, the level of expression in the F1 falls somewhere between the midparent value (complete additivity) and one of the parental values (complete parental allele dominance). When parental expression is not highly divergent, the mode of expression in the F1 appears to be that of overdominance.

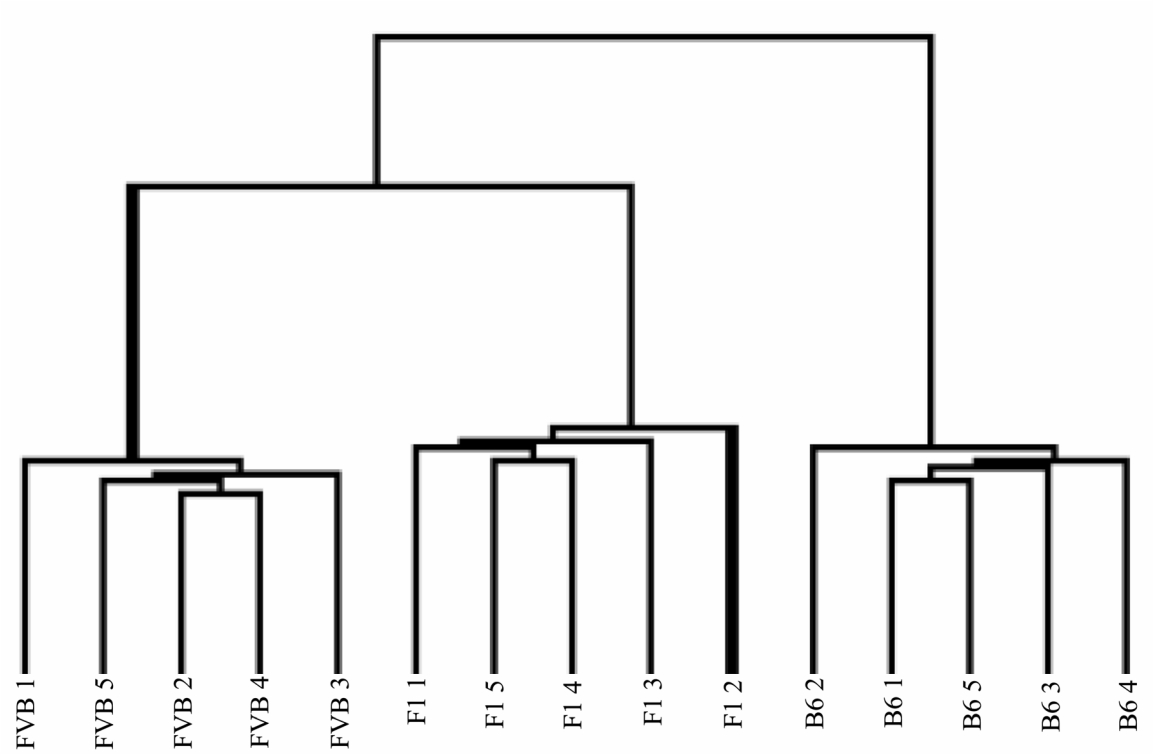


Figure 4.2: Clustering by genotype. Significantly divergent transcripts were clustered by experiment. Heirarchical clustering groups objects together based on the similarity between objects; in this case, gene expression. The length of the vertical arms in the clustering dendrogram is a measure of the linkage or correlation between groups. The shorter the arm, the stronger the correlation. The similarity metric used here was Pearson's correlation.

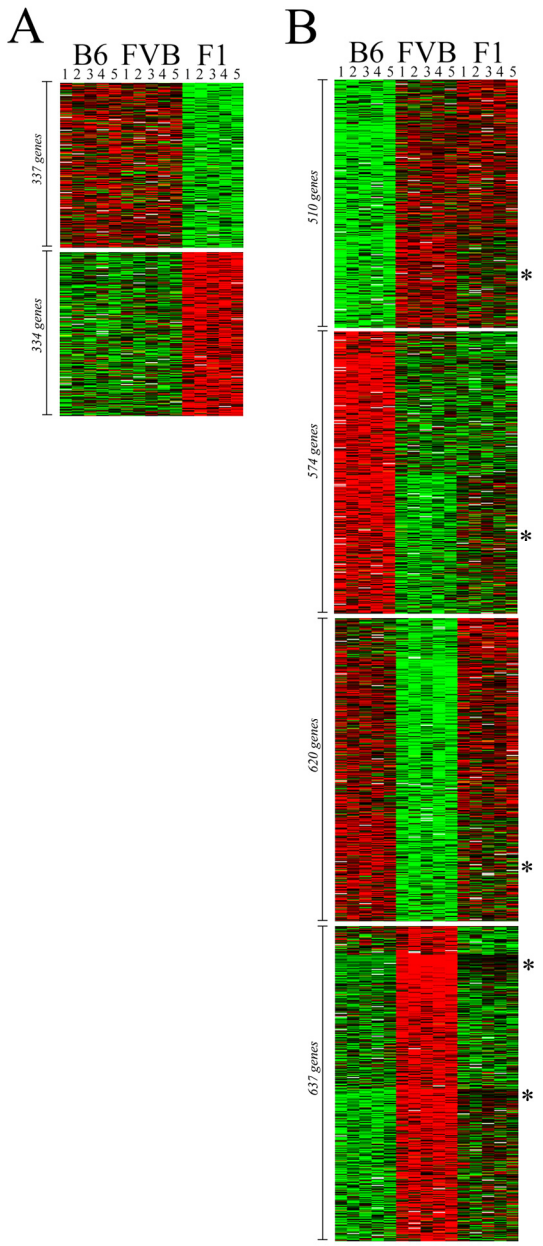


Figure 4.3: K-means clustering of the significantly divergent genes. Green represents upregulation and red represents downregulation. (A) Genes that show overdominant expression in the F1. The first cluster represents genes with enhanced expression compared to the parental strains and the second cluster represents genes with decreased expression compared to the parental strains. (B) Expression between parental strains is divergent in all four clusters. Expression in the F1 resembles expression in one parental strain. Asterix indicate examples of regions that appear to show expression in the F1 that is more additive than dominant.

Dominance effects are more prevalent in the F1 than additive effects

In order to evaluate additive and non-additive expression in a quantitative manner, statistical tests for additive, dominance and overdominance were applied to the entire dataset. Additive (a) and dominant (d) effects, along with the associated p-values, were calculated for each gene as described in the methods. When dominance effects are present, $|d|>0$ and the offspring deviate from the midparent value by the amount d. In the case that $d=0$, the inheritance pattern is completely additive. Due to biological and technical error, d is never completely 0 and significance testing is used to determine the effects of additivity.

An examination of the distribution of p-values revealed that there are many more transcripts that show very low p-values for dominance effects compared to additive effects (figure 4.4B and figure 4.4C). Assuming p-values are normally distributed, a large excess of low p-values indicates a true difference and the false discovery rate will be low. At the $p<0.01$ level, 842 transcripts are significant for the additive effect and 5,548 transcripts are significant for the dominance effect. When there is complete additivity, the value of d is close to 0 and the observed expression in the F1 hybrid is approximately equal to the midparent expression. If there is a strong correlation between the midparent value and the F1 hybrid then the expression in the F1 is influenced primarily by additivity. Many of the genes show a low correlation between the midparent value and F1 expression, especially at moderate intensity levels (figure 4.4A). Higher intensity levels are less abundant in the dataset but tend to show a better correlation, perhaps due to a ceiling effect in either expression biological expression levels or in fluorescence detection.

The ratio of the dominance effect to the additive effect (d/a) is another measure of non-additivity. Complete dominance can be visualized by $|d/a|=1$ while additivity occurs

when d/a=0. Any value of |d/a|>1 indicates overdominance. A plot of d/a reveals a large abundance of genes displaying nonadditive expression, although all forms of expression patterns are represented (figure 4.5). Once again, there were a greater number of genes that showed a dominance effect. It also appeared that there were more cases of B6 allele dominance then there were of FVB allele dominance.

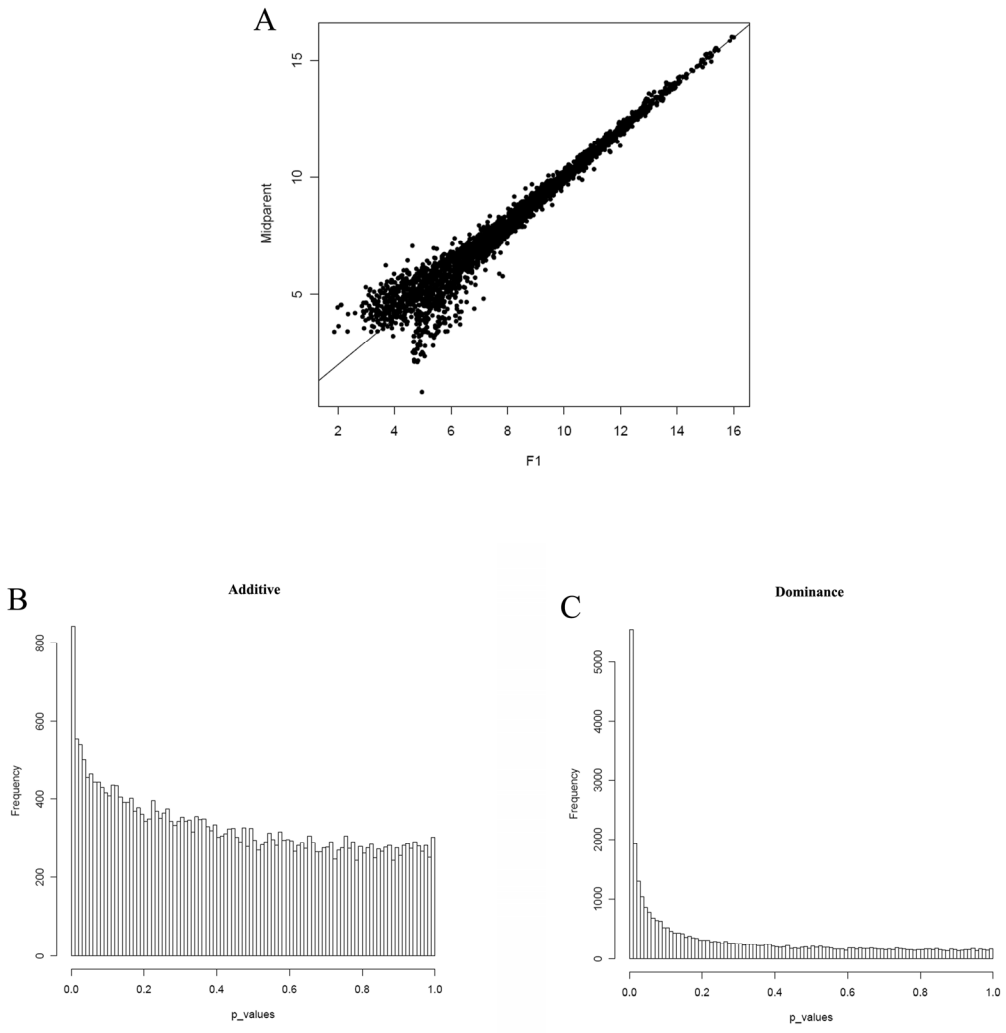


Figure 4.5: Analysis of additive and dominance effects. (Top) The F1 mean was plotted against the midparent mean. In the case of pure additivity the midparent value would equal the F1 mean and all the points would lie on the line with slope equal to 1. The lack of a strong correlation for much of the data indicates a strong dominance effect on F1 expression. (Bottom) Additive and dominance effects were tested for significance and the frequency of the resulting p-values were displayed as two histograms. Both effects have an excess of low p-values but there are far more significant results for dominance effects.

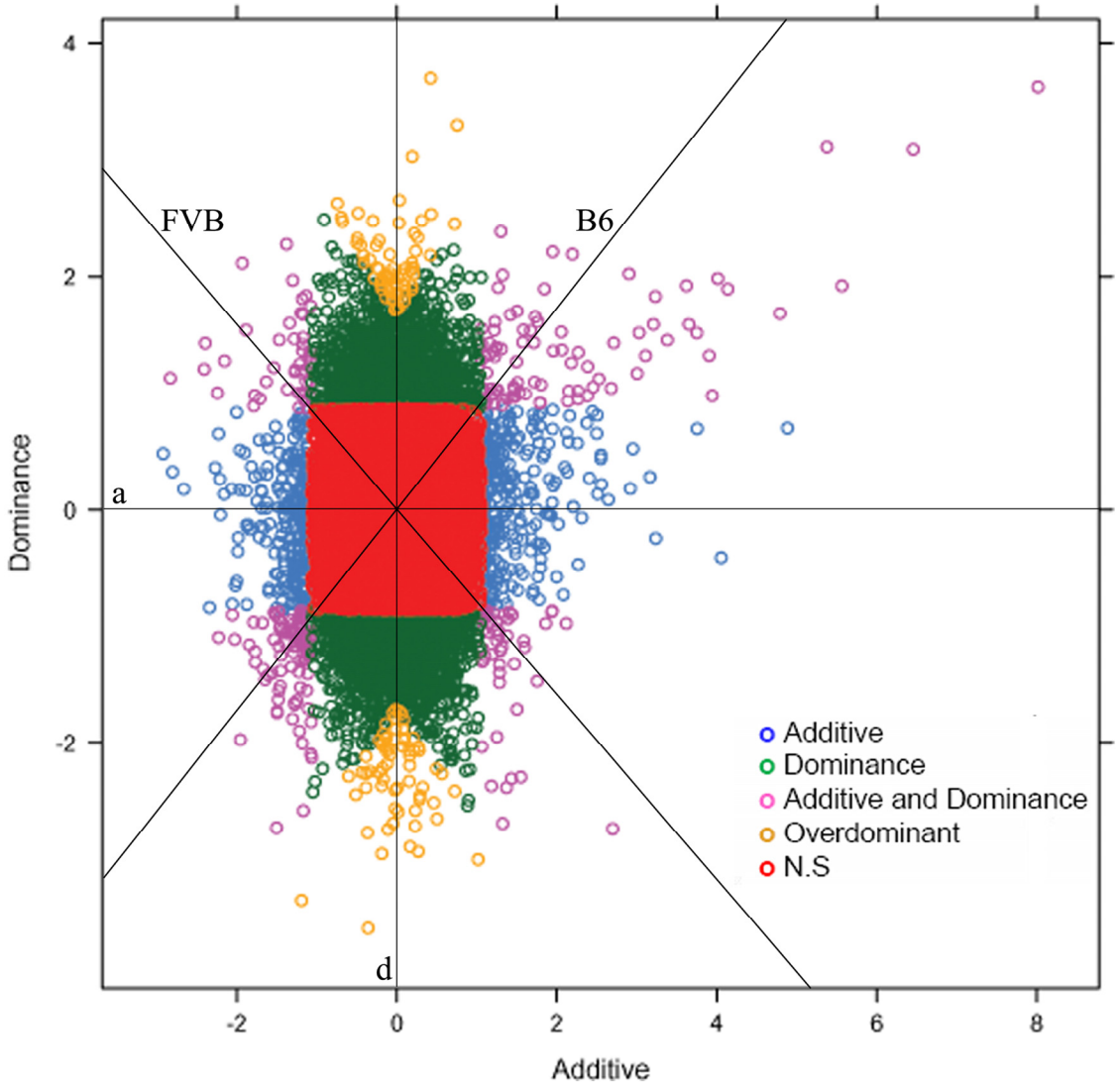


Figure 4.6: Scatterplot of additive and dominant expression patterns. The horizontal and vertical axis represent complete additivity and dominance respectively. The line with slope of 1 and -1 represents B6 and FVB allele dominance respectively. Red is not significant, green represents significant dominance effects and blue represents significant additive effects. Yellow represents significantly overdominant transcripts. Purple indicates significant dominance and additive effects, or allele dominance.

Analysis of the localization and function of significantly overdominant genes

Of the genes that were found to be significantly overdominant (OD), 65 were expressed above parental levels and 63 were expressed below parental levels (table 4.1). An analysis of function revealed that a wide range of cellular functions are represented and that many of the OD genes played a role in multiple cellular processes (table 4.2). Approximately 20% of the OD genes were membrane proteins and nearly 30% were involved in metabolism. Several OD genes have protein interaction domains (PH domains) and some were involved in transcriptional regulation and intracellular signaling processes. Interestingly, a large portion of the OD genes were of unknown function and were not annotated as known genes. Many of these transcripts lie within the boundaries of known genes and may actually represent alternatively spliced products that have not yet been officially associated with a gene.

Most OD genes appeared to be randomly distributed across the chromosomes (figure 4.6). Two chromosomes had the most abundant concentrations of OD genes; thirteen OD genes on chromosome 7 and nine OD genes on chromosome 11. No OD genes localized to chromosome 18. With the exception of chromosome 9, in which all OD genes were underdominant to the parental strains, both under and overdominant genes are represented on each chromosome. There are several instances were two or more OD genes appear to be co-localized near each other on chromosomes and in some of these cases the direction of expression is the same. In addition to mapping the location of the OD genes, the distribution of known single nucleotide polymorphisms (SNPs) between the parental strains was mapped onto each chromosome (figure 4.6). Many regions of the chromosomes contain dense SNP mapping, including several regions that contain OD genes. Approximately 50% of the OD genes contain SNPs between the parental strains (table 4.4). While many of the known SNPs were intronic or

synonomous within a coding region, 29 OD genes contained SNPs that were located within regions that could have an effect on transcription or translation: potential regulatory regions, untranslated regions (UTRs), or within a coding region and nonsynonomous (table 4.4). Fourteen OD genes actually contained SNPs in two or more of the previously mentioned regions.

Overdominance Genes (F1 over-expressed)

CLID	GBA	EntrezID	Name	Symbol	Chr	B6 AVG	FVB AVG	F1 AVG	B6vsFVB p (HSD)	B6vsFVB FC	F1vsB6p (HSD)	F1vsB 6 FC	F1vsFVB p (HSD)	F1vsFVB FC
H4065A11	BQ560560	545389	Centrosomal protein 170	Cep170	1	1.30	0.35	4.97	3.99E-01	-1.94	6.26E-04	-12.70	7.84E-05	-24.59
H3084B05	BG070141					1.79	2.93	5.07	3.03E-01	2.21	2.11E-03	-9.71	3.31E-02	-4.40
H4041C11	BQ556643					2.71	2.93	5.37	9.65E-01	1.17	2.40E-02	-6.30	3.79E-02	-5.41
H4048B11	BQ557774	57890	Interleukin 17 receptor E	Il17re	6	2.38	2.49	4.93	9.89E-01	1.08	1.50E-02	-5.84	1.94E-02	-5.41
H4047A08	BQ557586					3.44	3.43	5.87	1.00E+00	-1.01	3.22E-02	-5.38	3.14E-02	-5.43
614891	AI448892	14693	Guanine nucleotide binding protein, beta 2	Gnb2	5	3.96	4.02	6.30	9.96E-01	1.05	2.84E-02	-5.08	3.30E-02	-4.85
H4067D11	BQ560950	13730	Epithelial membrane protein 1	Emp1	6	3.85	4.34	6.15	6.29E-01	1.41	2.38E-03	-4.92	1.25E-02	-3.50
H3060A10	BG067951	11685	Arachidonate lipoxygenase, epidermal	Alox12e	11	4.54	4.25	6.82	9.38E-01	-1.22	4.41E-02	-4.88	2.40E-02	-5.95
H3037G06						3.90	4.31	6.17	8.26E-01	1.33	1.64E-02	-4.84	4.72E-02	-3.64
H4065B04	BQ560569	70699	Nucleoporin 205	Nup205	6	2.60	2.76	4.81	9.69E-01	1.12	1.77E-02	-4.63	2.72E-02	-4.14
H3079E05	BG069732	18973	Polymerase (DNA directed), epsilon	Pole	5	4.16	4.37	6.37	9.47E-01	1.16	1.57E-02	-4.60	2.76E-02	-3.98
H4028D10	BQ554494	66580	ESF1, nucleolar pre-rRNA processing protein, homolog (S. cerevisiae)	Esf1	2	3.48	4.05	5.59	5.52E-01	1.49	5.32E-03	-4.33	3.58E-02	-2.91
H3082G11	BG069939	328451	Predicted gene, EG328451	EG328451		3.71	3.87	5.80	9.60E-01	1.12	9.57E-03	-4.25	1.56E-02	-3.81
AI838755	AI838755	319939	Tensin 3	Tns3	11	3.92	3.43	5.97	7.80E-01	-1.40	3.57E-02	-4.14	1.06E-02	-5.80
621079	AI451037	28193	Receptor accessory protein 3	Reep3		3.80	4.16	5.77	8.12E-01	1.29	1.45E-02	-3.91	4.40E-02	-3.04
AI844823	AI844823	269587	Erythrocyte protein band 4.1	Epb4.1	4	5.81	5.93	7.71	9.81E-01	1.09	2.93E-02	-3.73	4.08E-02	-3.44
H4034H06	BQ555581					3.86	2.89	5.51	8.06E-02	-1.97	4.46E-03	-3.13	9.15E-05	-6.16
H3118F01	BG073090	29870	G two S phase expressed protein 1	Gtse1	15	3.93	3.76	5.54	9.48E-01	-1.13	3.41E-02	-3.04	1.94E-02	-3.43
H3010A05						4.67	4.36	6.20	8.24E-01	-1.24	2.86E-02	-2.89	9.75E-03	-3.58
H3064F06	BG068361					3.54	3.27	5.01	8.53E-01	-1.21	2.92E-02	-2.78	1.11E-02	-3.36
H3113C12						4.15	4.06	5.34	9.71E-01	-1.06	2.39E-02	-2.28	1.58E-02	-2.43
AI853640	AI853640		Transcribed locus			5.01	4.72	6.15	6.91E-01	-1.22	1.71E-02	-2.21	3.87E-03	-2.70
H3046H05	BG079949	170719	Oxidation resistance 1	Oxr1	15	5.30	5.27	6.27	9.93E-01	-1.02	3.65E-03	-1.96	3.00E-03	-1.99
H4069F11	BQ561335	81535	Sphingosine-1-phosphate phosphatase 1	Sgpp1	12	5.57	5.87	6.50	4.15E 01	1.23	3.89E 03	1.91	4.05E 02	1.55
H3090G09	BG070754	22632	YY1 transcription factor	Yy1	12	5.82	6.00	6.74	7.69E-01	1.13	9.25E-03	-1.89	3.23E-02	-1.67
H3120H02						5.08	5.26	5.98	7.10E-01	1.13	4.82E-03	-1.86	2.02E-02	-1.64
H3001H08	BG063079	11535	Adrenomedullin	Adm	7	5.00	5.16	5.87	8.24E-01	1.11	1.31E-02	-1.82	3.83E-02	-1.64
H4058C06	BQ559442	72129	Peroxisomal biogenesis factor 13	Pex13	11	5.94	5.93	6.74	1.00E+00	-1.00	2.61E-02	-1.75	2.51E-02	-1.75
H3009A08	BG063582	56194	PRP40 pre-mRNA processing factor 40 homolog A (yeast)	Prpf40a	2	4.82	4.66	5.61	8.23E-01	-1.12	2.56E-02	-1.74	8.73E-03	-1.94
AI854740	AI854740	11800	Apoptosis inhibitor 5	Api5	2	5.93	6.18	6.72	3.16E-01	1.19	1.21E-03	-1.73	1.75E-02	-1.45
H3002C06	BG076418	68626	ElaC homolog 2 (E. coli)	Elac2	11	5.37	5.36	6.14	9.98E-01	-1.01	2.29E-02	-1.71	2.06E-02	-1.72
H4070C09	BQ561431	240756	Kelch-like 12 (Drosophila)	Klhl12		4.99	5.02	5.74	9.93E-01	1.02	1.94E-02	-1.67	2.37E-02	-1.64
H3046A07	BG066705					4.57	4.23	5.28	4.07E-01	-1.27	4.07E-02	-1.64	3.80E-03	-2.08
AI846740	AI846740	13481	Dolichol-phosphate (beta-D) mannosyltransferase 2	Dpm2	2	8.57	8.68	9.23	7.68E-01	1.08	4.67E-03	-1.58	1.63E-02	-1.46
H4009F07	BQ551495	270210	Zinc finger protein 651	Zfp651		6.53	6.44	7.19	8.13E-01	-1.07	2.02E-03	-1.58	7.14E-04	-1.68
H3022A03	BG064622		Adult male spinal cord cDNA, RIKEN full-length enriched library, clone:A330053K08 product:unclassifiable, full insert sequence			6.89	6.62	7.53	3.87E-01	-1.20	1.57E-02	-1.56	1.41E-03	-1.87
AI854533	AI854533	320973	RIKEN cDNA D330023K18 gene	D330023K18Rik		6.87	6.35	7.50	9.59E-02	-1.44	4.50E-02	-1.54	8.10E-04	-2.21
AI836222	AI836222	20333	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	Sec22b	3	8.93	9.19	9.52	8.84E-02	1.20	4.95E-04	-1.51	2.87E-02	-1.26
H3077C10	BG069525	50505	Excision repair cross-complementing rodent repair deficiency, complementation group 4	Ercc4	16	7.26	7.45	7.84	3.61E-01	1.14	2.78E-03	-1.49	3.50E-02	-1.30
H3153H11	BG075968	381546	Coiled-coil domain containing 24	Ccdc24		6.49	6.56	7.06	9.08E-01	1.05	1.42E-02	-1.48	3.03E-02	-1.41
H3121C09	BG073323	67772	Chromodomain helicase DNA binding protein 8	Chd8	14	6.10	5.97	6.66	8.26E-01	-1.09	4.37E-02	-1.48	1.51E-02	-1.61
AI843285	AI843285		Transcribed locus			7.64	7.74	8.18	8.07E-01	1.07	1.32E-02	-1.45	4.09E-02	-1.35
UTDDEST110527						5.70	5.54	6.21	5.92E-01	-1.11	1.31E-02	-1.43	2.24E-03	-1.59
AI843257	AI843257	227622	CDNA sequence BC029214	BC029214	2	7.37	7.68	8.08	7.26E-01	1.08	8.42E-03	-1.43	3.37E-02	-1.32
H3007F08	CK334174	58998	Poliovirus receptor-related 3	Pvrl3	16	6.82	6.76	7.30	9.33E-01	-1.04	4.04E-02	-1.39	2.14E-02	-1.45
AI839645	AI839645	18813	Proliferation-associated 2G4	Pa2g4	10	8.00	8.11	8.46	6.61E-01	1.08	9.79E-03	-1.38	4.75E-02	-1.27
AI836628	AI836628	216821	Transmembrane protein 11	Tmem11	11	10.32	10.47	10.78	4.27E-01	1.11	4.70E-03	-1.38	4.71E-02	-1.24
H3094B06	BG071049	20658	Son cell proliferation protein	Son	16	8.37	8.27	8.82	7.89E-01	-1.07	3.03E-02	-1.36	9.26E-03	-1.46
AI837635	AI837635	66437	Fission 1 (mitochondrial outer membrane) homolog (yeast)	Fis1	5	7.16	7.17	7.60	9.97E-01	1.01	4.35E-02	-1.36	4.94E-02	-1.34
H3028D05	BG065197	19983	Ribosomal protein L5	Rpl5	7	10.30	10.25	10.71	9.38E-01	-1.03	2.33E-02	-1.33	1.26E-02	-1.37
H4054F07	BQ558844					7.30	7.31	7.70	9.97E-01	1.01	4.31E-02	-1.31	4.86E-02	-1.30
368696	AI415258	17172	Achaete-scute complex homolog-like 1 (Drosophila)	Ascl1	10	7.08	6.99	7.47	8.23E-01	-1.06	4.83E-02	-1.31	1.66E-02	-1.39
H3042C07	BG066377	233532	Remodeling and spacing factor 1	Rsf1	7	8.76	8.69	9.12	7.65E-01	-1.05	1.37E-02	-1.28	3.91E-03	-1.35
H3079C08	BG069711	56224	Tetraspanin 5	Tspan5	3	8.80	8.83	9.15	9.79E-01	1.02	3.03E-02	-1.27	4.30E-02	-1.25
721798	AI661184	26908	Eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked	Eif2s3y	Y	9.03	9.02	9.37	9.95E-01	-1.01	6.20E-03	-1.26	5.22E-03	-1.27
H3040G07	BG066254	11750	Annexin A7	Anxa7	14	6.88	6.84	7.21	9.47E-01	-1.03	4.91E-02	-1.26	2.80E-02	-1.29
445988	AI324156	217365	Nuclear protein localization 4 homolog (S. cerevisiae)	Nploc4	11	7.18	7.18	7.50	1.00E+00	-1.00	2.61E-02	-1.25	2.59E-02	-1.25
H3086E01	BG070355		Transcribed locus			8.35	8.47	8.64	1.22E-01	1.09	8.66E-04	-1.22	3.74E-02	-1.12
H3156A10	CK335251	13033	Cathepsin D	Ctsd	7	8.91	8.90	9.17	9.95E-01	-1.01	4.78E-02	-1.20	4.03E-02	-1.21
AI835654	AI835654	66425	Purkinje cell protein 4 like 1	Pep4l1	1	9.73	9.74	9.99	1.00E+00	1.00	4.55E 02	1.19	4.57E 02	1.19
AI839261	AI839261	68066	Solute carrier family 25, member 39	Slc25a39	11	10.03	10.06	10.27	9.09E-01	1.02	2.37E-02	-1.18	4.99E-02	-1.15
H3126H03	BG073680	216150	Cell division cycle 34 homolog (S. cerevisiae)	Cdc34	10	10.44	10.52	10.66	2.93E-01	1.06	2.30E-03	-1.17	3.78E-02	-1.11
AI837752	AI837752	56177	Olfactomedin 1	Olfm1	2	11.55	11.57	11.77	9.71E-01	1.01	1.56E-02	-1.16	2.37E-02	-1.15
AI851981	AI851981	66092	Growth hormone inducible transmembrane protein	Ghitm	14	10.43	10.39	10.60	7.38E-01	-1.03	2.21E-02	-1.13	5.75E-03	-1.16
AI837422	AI837422	22154	Tubulin, beta 5	Tubb5	17	12.73	12.80	12.88	1.14E-01	1.04	7.64E-04	-1.11	3.51E-02	-1.06

Table 4.1: Significant overdominant genes; enhanced in F1. All transcripts that show enhanced expression in the F1 compared to either parental strain are shown. CLID, clone identification; GBA, genebank accession; EntrezID, Entrez gene number, Chr, chromosome; AVG, average, HSD, Tukey's post hoc test *p*-value; FC, fold change; blue, B6; green, F1; yellow, FVB. A negative number indicates the mean of the first strain is higher than the mean of the second strain and a positive number indicates the opposite. Criterion for significance is $|d/a|>1$ and a significant difference ($p<0.05$) in expression of transcript in the F1 compared to each parental strain.

Overdominance Genes (F1 under-expressed)

CLID	GBA	EntrezID	Name	Symbol	Chr	B6 AVG	FVB AVG	F1 AVG	B6vsFVB p (HSD)	B6vsFVB FC	F1vsB6p (HSD)	F1vsB 6 FC	F1vsFVB p (HSD)	F1vsFVB FC
421708	AI429747	208638	Solute carrier family 25, member 38	Slc25a38	9	4.31	4.80	2.10	5.17E-01	1.40	6.89E-04	4.62	1.18E-04	6.46
573163	AI428933	77531	Ankyrin repeat and sterile alpha motif domain containing 1B	Anks1b	10	5.28	4.45	3.19	2.14E-01	-1.79	2.01E-03	4.27	4.83E-02	2.39
H3054C04	BG080487					5.88	5.95	4.51	9.89E-01	1.05	4.84E-02	2.57	3.76E-02	2.70
355136	AI414955	13875	Ets2 repressor factor	Erf	7	6.40	5.80	5.11	9.54E-02	-1.51	8.82E-04	2.45	4.96E-02	1.62
AI847288	AI847288	383103	CDNA sequence BC068110	BC068110	16	6.18	6.03	5.11	8.77E-01	-1.11	1.08E-02	2.09	2.61E-02	1.89
H3110F07	BG072450					4.96	4.77	3.92	7.95E-01	-1.15	1.29E-02	2.06	4.15E-02	1.80
642805	AI451587					6.44	6.20	5.43	6.56E-01	-1.18	7.65E-03	2.02	3.77E-02	1.71
AI850724	AI850724	71955	RIKEN cDNA 2400003C14 gene	2400003C14Rik	8	4.96	6.07	3.94	7.10E-03	2.16	1.30E-02	2.02	3.08E-05	4.36
AI839946	AI839946	66624	Signal peptidase complex subunit 2 homolog (S. cerevisiae)	Spcs2	7	5.89	6.03	4.89	9.04E-01	1.10	2.78E-02	1.99	1.28E-02	2.20
H4059E07	BQ559645	76815	Calcium binding and coiled-coil domain 2	Calcoco2	11	4.77	4.68	3.81	9.46E-01	-1.07	1.65E-02	1.95	2.93E-02	1.83
AI838676	AI838676	26442	Proteasome (prosome, macropain) subunit, alpha type 5	Psma5	3	9.71	9.68	8.87	9.95E-01	-1.02	4.09E-02	1.80	4.85E-02	1.76
H3015G04	BG064122	66455	RIKEN cDNA 2610019P18 gene	2610019P18Rik	5	5.82	6.02	4.99	7.77E-01	1.15	4.08E-02	1.77	1.20E-02	2.03
551860	AI429349	192232	Hermansky-Pudlak syndrome 4 homolog (human)	Hps4	5	5.88	6.15	5.12	5.57E-01	1.20	2.66E-02	1.70	4.02E-03	2.04
H3052H05	BG067304	98732	RAB3 GTPase activating protein subunit 2	Rab3gap2	1	5.76	5.93	5.03	7.62E-01	1.12	1.89E-02	1.66	5.31E-03	1.86
AI843847	AI843847	17977	Nuclear receptor coactivator 1	Ncoa1	12	7.76	7.71	7.07	9.79E-01	-1.03	3.53E-02	1.62	5.00E-02	1.56
AI843343	AI843343	1E+08	Hypothetical protein LOC100040913	LOC100040913	1	8.87	8.82	8.22	9.67E-01	-1.04	2.34E-02	1.57	3.65E-02	1.52
AI835817	AI835817	60599	Transformation related protein 53 inducible nuclear protein 1	Trp53inp1	4	6.51	6.44	5.87	8.56E-01	-1.05	1.09E-03	1.56	2.72E-03	1.48
AI854820	AI854820	19356	RAD17 homolog (S. pombe)	Rad17	13	6.65	6.66	6.01	9.98E-01	1.01	4.44E-03	1.56	4.02E-03	1.57
H4051E07	BQ558326	101490	Inositol polyphosphate-5-phosphatase F	Inpp5f	7	6.06	6.18	5.43	8.30E-01	1.09	3.32E-02	1.54	1.16E-02	1.68
AI850133	AI850133	223864	Rap guanine nucleotide exchange factor (GEF) 3	Rapgef3	15	7.78	7.55	7.17	2.95E-01	-1.17	2.72E-03	1.52	4.46E-02	1.31
H3112G10	BG072606	13132	Disabled homolog 2 (Drosophila)	Dab2	15	6.20	6.21	5.61	9.98E-01	1.01	6.92E-03	1.50	6.19E-03	1.51
583640	AI450544	56790	DNA segment, Chr 3, ERATO Doi 300, expressed	D3Ertd300e		6.29	6.39	5.74	8.62E-01	1.07	4.09E-02	1.46	1.61E-02	1.57
H3002B08						6.35	6.39	5.82	9.39E-01	1.03	3.46E-03	1.44	1.93E-03	1.49
537177	AI426951	66917	Cysteine and histidine-rich domain (CHORD)-containing, zinc-binding protein 1	Chordc1	9	9.57	9.47	9.07	6.41E-01	-1.08	2.37E-03	1.42	1.19E-02	1.32
H3037H11	BG066015	68059	Transmembrane 9 superfamily member 2	Tm9sf2	14	9.88	9.81	9.43	7.44E-01	-1.05	2.11E-03	1.37	7.76E-03	1.30
AI853828	AI853828	13803	Ectodermal-neural cortex 1	Enc1	13	7.82	7.88	7.38	9.12E-01	1.04	2.45E-02	1.37	1.18E-02	1.42
H4017A11	BQ552623	67846	Transmembrane protein 39a	Tmem39a	16	7.08	7.14	6.64	9.14E-01	1.04	3.32E-02	1.35	1.60E-02	1.41
H3140F06	BG074894	71853	Protein disulfide isomerase associated 6	Pdia6	12	9.64	9.55	9.24	6.38E-01	-1.06	2.55E-03	1.33	1.31E-02	1.25
AI851128	AI851128	14733	Glypican 1	Gpc1	1	7.47	7.63	7.07	4.68E-01	1.12	2.91E-02	1.32	3.34E-03	1.48
H3081A07	BG069871	77644	RIKEN cDNA C330007P06 gene	C330007P06Rik	X	7.26	7.26	6.87	9.98E-01	-1.00	3.69E-03	1.31	4.13E-03	1.31
574540	AI448632	233033	Sterile alpha motif domain containing 4B	Samd4b	7	6.99	6.90	6.62	6.68E-01	-1.06	8.50E-03	1.30	4.04E-02	1.22
H3121G02	BG073362	68035	RIKEN cDNA 3100004P22 gene	3100004P22Rik	7	9.76	9.64	9.38	3.39E-01	-1.08	1.36E-03	1.29	1.80E-02	1.19
AI849943	AI849943		Transcribed locus			10.17	10.14	9.80	9.47E-01	-1.03	1.73E-02	1.29	3.05E-02	1.26
H4022A05	BQ553434	66871	Copine VIII	Cpne8	15	7.78	7.85	7.41	8.30E-01	1.04	1.23E-02	1.29	4.32E-03	1.35
AI854768	AI854768	75302	Additional sex combs like 2 (Drosophila)	Asxl2	12	8.98	8.95	8.62	9.67E-01	-1.02	1.50E-02	1.28	2.35E-02	1.26
H4048E11	BQ557838	217149	Melanoma nuclear protein 13	Mel13		9.38	9.34	9.03	8.74E-01	-1.03	5.21E-03	1.28	1.27E-02	1.24
H3017B11	BG064234	53328	Progesterone receptor membrane component 1	Pgrmc1	X	10.93	10.75	10.58	1.86E-02	-1.13	9.96E-05	1.27	2.18E-02	1.13
H3143B04	BG075084	12955	Crystallin, alpha B	Cryab	9	11.92	11.87	11.57	8.64E-01	-1.03	7.45E-03	1.27	1.89E-02	1.23
H3140F11	BG074899	235380	Dmx-like 2	Dmxl2	9	12.75	12.75	12.41	9.99E-01	1.00	4.76E-02	1.27	4.48E-02	1.27
AI849056	AI849056	108911	Regulator of chromosome condensation 2	Rcc2	4	8.63	8.76	8.30	4.71E-01	1.09	2.06E-02	1.26	2.41E-03	1.38
H4034F01	BQ555529	330695	Cortexin 1	Ctxn1	8	13.26	13.21	12.94	7.51E-01	-1.04	1.49E-03	1.25	5.31E-03	1.21
AI848774	AI848774	214968	Sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	Sema6d	2	8.70	8.67	8.38	9.59E-01	-1.02	2.85E-02	1.25	4.65E-02	1.22
AI847663	AI847663	54484	Makorin, ring finger protein, 1	Mkrn1	6	10.50	10.44	10.18	6.78E-01	-1.04	7.51E-04	1.25	3.21E-03	1.20
AI836943	AI836943	22608	Y box protein 1	Ybx1	4	10.64	10.66	10.35	9.64E-01	1.02	1.94E-02	1.23	1.21E-02	1.25
AI849425	AI849425	14402	Gamma-aminobutyric acid (GABA-A) receptor, subunit beta 3	Gabrb3	7	10.73	10.63	10.44	2.84E-01	-1.07	8.48E-04	1.22	1.37E-02	1.14
H3133C01	BG074285	20918	Eukaryotic translation initiation factor 1	Eif1	11	13.85	13.80	13.58	7.74E-01	-1.04	9.16E-03	1.21	3.16E-02	1.17
AI852332	AI852332	76932	ADP-ribosylation factor interacting protein 2	Arfip2	7	9.28	9.32	9.02	8.57E-01	1.03	1.37E-02	1.20	5.28E-03	1.23
H3120G02	BG073274	14678	Guanine nucleotide binding protein, alpha inhibiting 2	Gnai2	9	10.33	10.34	10.09	9.95E-01	1.01	2.45E-02	1.18	2.08E-02	1.19
H3155F03	CK335248	20085	Ribosomal protein S19	Rps19	7	12.29	12.32	12.04	8.30E-01	1.02	2.42E-03	1.18	8.95E-04	1.21
H3099H01						11.08	11.15	10.84	5.05E-01	1.05	6.53E-03	1.18	9.02E-04	1.24
H3074A11	BG069222	218699	PX domain containing serine/threonine kinase	Pxk	14	11.81	11.81	11.61	1.00E+00	-1.00	6.75E-03	1.15	7.03E-03	1.15
H3115H01	BG072863	268449	Ribosomal protein L23a	Rpl23a	14	12.83	12.83	12.63	9.99E-01	1.00	2.96E-02	1.15	2.78E-02	1.15
AI838282	AI838282		Transcribed locus, strongly similar to XP_001104203.1 similar to tubulin, beta 8 isoform 2 [Macaca mulatta]			13.67	13.69	13.48	9.58E-01	1.01	2.25E-02	1.14	1.36E-02	1.15
AI845039	AI845039	269642	N-acetyltransferase 8-like	Nat8l	5	13.42	13.44	13.24	9.27E-01	1.01	1.72E-02	1.14	8.85E-03	1.15
H4022G02	BQ553543	67298	G protein-coupled receptor associated sorting protein 1	Gprasp1	X	11.84	11.94	11.67	2.78E-01	1.07	2.48E-02	1.13	1.44E-03	1.21
AI850479	AI850479	27360	Adducin 3 (gamma)	Add3	19	9.32	9.32	9.15	9.97E-01	1.00	4.04E-02	1.12	3.57E-02	1.13
H3048C07						11.58	11.65	11.43	3.85E-01	1.05	3.44E-02	1.11	2.97E-03	1.17
H3122E12	BG073436	11947	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	Atp5b	10	10.76	10.78	10.61	8.48E-01	1.02	1.82E-02	1.11	6.79E-03	1.13
AI846609	AI846609	11674	Aldolase 1, A isoform	Aldoa	7	12.35	12.38	12.21	9.06E-01	1.02	4.76E-02	1.11	2.24E-02	1.12
617269	AI450671	238505	5-methyltetrahydrofolate-homocysteine methyltransferase	Mtr	13	13.21	13.22	13.07	9.89E-01	1.00	3.36E-03	1.11	2.61E-03	1.11
H3149E05	BG075626	11606	Angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	Agt	8	12.83	12.81	12.69	8.70E-01	-1.01	9.53E-03	1.10	2.37E-02	1.09
AI837950	AI837950					11.94	12.02	11.83	1.81E-01	1.06	4.99E-02	1.08	1.72E-03	1.14
H3104H04	CK334754		Transcribed locus			12.01	12.05	11.90	4.67E-01	1.03	4.38E-02	1.08	4.99E-03	1.11

Table 4.2: Significant overdominant genes; reduced in F1. All transcripts that show a reduction in expression in the F1 compared to either parental strain are shown. CLID, clone identification; GBA, genebank accession; EntrezID, Entrez gene number, Chr, chromosome; AVG, average, HSD, Tukey's post hoc test *p*-value; FC, fold change; blue, B6; green, F1; yellow, FVB. A negative number indicates the mean of the first strain is higher than the mean of the second strain and a positive number indicates the opposite. Criterion for significance is |d/a|>1 and a significant difference ($p<0.05$) in expression of transcript in the F1 compared to each parental strain.

DAVID and WebGestalt Analysis of Functional Groups for Overdominant Genes

Location	Genes	
Integral to membrane (CC)	*Tm9sf2* , **Emp1**, **Tspan5**, *Gabrb3* , **Fis1**, *BC068110* , *Gprasp1* , **Tmem11**, **Pvrl3**, **Slc25a39**, *Sema6d* , **Ghitm, Pex13**, **Sgpp1**, *Slc25a38* , *Pgrmc1* , **Sec22b**, *Atp5b*, *Ctxn1* , *Spcs2* , **Nup205, *Tmem39a*** , **Dpm2**	
Nuclear envelope-ER network (CC)	**Sec22b**, *Spcs2* , **Dpm2**	*
Endoplasmic reticulum (CC)	*Pdia6* , Sgpp1, *Pgrmc1* , **Sec22b**, **Nploc4**, *Spcs2* , **Olfm1**, **Dpm2**	*
Cytoskeleton (CC)	*Enc1* , **Sgpp1**, **Tubb5**, **Addc3**, *Rcc2* , **Gtse1**, **Epb4.1**	
Microtubule (CC)	**Tubb5**, *Rcc2* , **Gtse1**	
Ribosome (CC)	*Rpl23a* , *Rps19* , **Rpl5**	
Mitochondrion (CC)	**Slc25a39**, **Sgpp1**, *Slc25a38* , **Fis1**, **Ctsd**, *Atp5b*	
Activity	**Genes**	
Actin binding (MF)	*Enc1* , **Pxk**, **Epb4.1**	
Peptidase activity (MF)	*Enc1* , **Ctsd**, *Psma5* , **Pa2g4**, *Erf*, *Spcs2*	
GTP binding (MF)	*Arfip2* , **Tubb5**, **Eif2s3y**, *Gnai2*	
GTPase activity (MF)	**Gnb2**, **Tubb5**, *Gnai2*	
Domain	**Genes**	
Phosphotyrosine interaction region(Interpro)	**Tns3**, *Dab2* , *Anks1b*	*
Pleckstrin homology-type (Interpro)	**Tns3**, *Dab2* , *Anks1b* , **Epb4.1**	
Function	**Genes**	
Transcription		
Regulation of Transcription (BP)	**Rsf1**, *Ncoa1* , **Ascl1**, **Pa2g4**, *Erf*, *Ybx1* , **Yy1**	
Transcription factor activity (MF)	**Pa2g4**, *Erf*, **Yy1**	
Repressor (SP)	*Erf*, *Ybx1* , **Yy1**	
Chromatin binding		
Chromatin binding (MF)	*Ncoa1* , *Mkrn1* , **Chd8**	*
Signalling		
Signal transduction (BP)	*Arfip2* , *Ncoa1* , *Cryab* , **Tns3**, *Gabrb3* , **Gnb2**, *Agt* , *Rapgef3* , **Ascl1**, *Pxk* , **Tmem11,** *Gnai2*	
Cell surface receptor linked signal transduction (BP)	*Cryab* , *Gabrb3* , **Gnb2**, **Ascl1**, *Gnai2*	
Intracellular signaling cascade (BP)	*Arfip2* , **Tns3**, *Agt* , *Rapgef3* , *Pxk* , *Gnai2*	
G-protein coupled receptor signalling pathway (BP)	*Gabrb3* , **Gnb2**, *Gnai2*	
Transport and Localization		
Localization (BP)	*Arfip2* , *Tm9sf2* , *Pdia6* , **Tspan5**, *Gabrb3* , **Pex13**, **Tubb5,** *Dab2* , **Sec22b**, *Atp5b* , **Ascl1**, **Alox12e,** *Gprasp1* , *Dmxl2* , *Rad17* , **Slc25a39**, *Slc25a38* , **Epb4.1**, *Hps4*	
Transport (BP)	*Tm9sf2* , *Pdia6* , *Gabrb3* , **Tubb5**, *Dab2* , **Alox12e**, *Gprasp1* , *Dmxl2* , *Rad17* , **Slc25a39**, **Pex13**, *Slc25a38* , **Sec22b**, *Atp5b* , *Hps4*	
Apoptosis		
Apoptosis (BP)	*Trp53inp1* , **Sgpp1**, Api5, *Agt* , **Fis1**	
Regulation of apoptosis(BP)	*Trp53inp1* ,**Api5**, *Agt*	
Stress response		
Response to stress (BP)	*Cryab* , *Trp53inp1* , *Agt* , **Pole**, **Ercc4**, *Rad17*	
Response to DNA damage stimulus (BP)	**Pole**, **Ercc4**, *Rad17*	
Metabolism		
Metabolism (BP)	**Oxr1**, **Cdc34**, **Prpf40a**, *Rapgef3* , *Psma5* , *Rps19* , *Pxk* , **Ercc4**, *Aldoa* , **Rpl5**, **Tns3**, **Elac2**, **Sgpp1**, *Enc1* , **Nploc4, Pole**, *Spcs2* , **Dpm2**, *Ncoa1* , *Rpl23a* , *Pdia6* , **Tubb5**, **Ascl1**, *Agt* , **Alox12e**, *Mtr* , **Pa2g4**, *Rad17* , **Rsf1**, *Eif1* , **Ctsd**, **Eif2s3y**, *Atp5b* , *Ybx1* , *Erf*, **Olfm1**, **Yy1**, *Hps4*	
Fatty acid biosynthesis (BP)	*Agt* , **Alox12e**	*
Protein Metabolism (BP)	*Rpl23a* , **Cdc34**, **Tubb5**, *Rapgef3* , *Psma5* , **Pa2g4**, *Pxk* , *Rps19* , **Rpl5**, **Tns3**, *Enc1* , *Eif1* , **Ctsd, Eif2s3y, Nploc4,** *Ybx1* , *Erf*, **Olfm1**, *Spcs2* , *Hps4*	
Catalytic activity (MF)	**Cdc34**, *Psma5* , *Pxk* , **Ercc4**, *Aldoa* , **Tns3**, *Nat8l* , **Sgpp1**, *Enc1* , **Elac2**, **Pole**, *Spcs2* , **Dpm2, Ncoa1**, **Pdia5, Tubb5**, **Alox12e**, **Mtr**, **Pa2g4**, **Rad17**	
Development		
Cell diferentiation (BP)	*Sema6d* , *Dab2* , *Agt* , **Ascl1**, *Hps4*	
Nervous System Development (BP)	*Sema6d* , *Agt* , **Ascl1**	

Table 4.3: Function of overdominant genes. DAVID and WebGestalt were used for functional group analysis. Function is described in the left column. The gene symbols are shown in the middle column. Bold lettering indicates overdominance and italics indicate underdominance. Asterisks indicate that a fuctional group was determined to be significantly overrepresented (p<0.01) by either DAVID or WebGestalt (see methods).

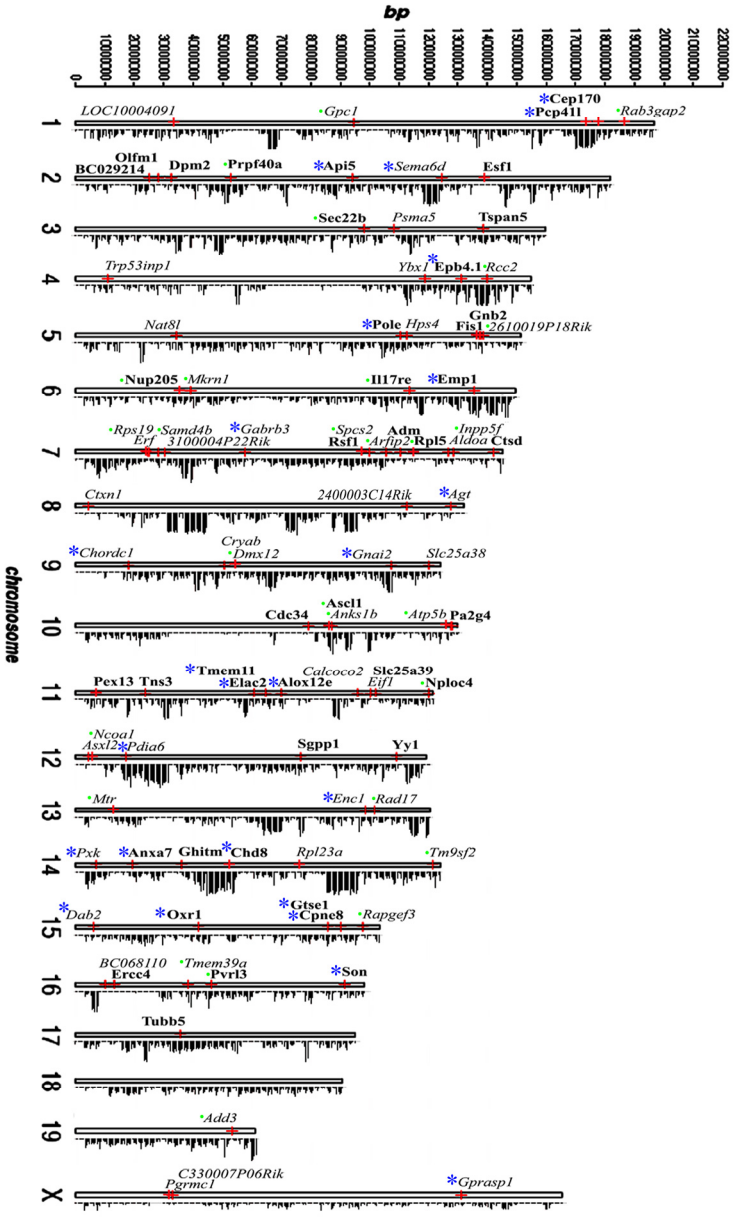


Figure 4.7: Distribution of overdominant genes and SNPs. Approximate gene location is indicated by the red cross. Gene symbol shown to the left of the chromosome and relative SNP density shown to the right. Italics, underdominance; bold,overdominance; green dots, 1 or more SNPs in intronic regions or withing coding regions that are predicted to be synomous (a synomous mutation does not result in a change in the amino acid sequence); blue asterisks, SNP is nonsynomous, within regulatory region or in UTR.

Gene	I	L	UTR	Cs	Cn	Undefined	Total
Add3	0	0	0	0	0	1	1
Agt	9	1	4	0	4	0	18
Alox12e	0	1	0	0	1	0	2
Anks1b	427	0	0	0	0	57	484
Anxa7	2	1	3	0	1	0	7
Api5	10	0	1	0	0	1	12
Sec22b	1	0	0	0	0	0	1
Arfip2	1	0	0	0	0	0	1
Ascl1	0	0	0	0	0	1	1
Atp5b	1	0	0	0	0	0	1
Cep170	20	1	12	3	1	4	41
Chd8	13	2	0	6	1	25	47
Chordc1	0	1	0	0	0	0	1
Cpne8	54	1	3	0	0	3	61
D3Ertd300e	1	0	1	0	0	0	2
Dab2	0	1	102	0	0	1	104
Dnxl2	0	0	0	0	0	1	1
Eif2s3y	0	0	0	0	0	51	51
Elac2	2	1	0	0	0		3
Emp1	8	2	0	0	0	0	10
Enc1	0	0	0	0	0	8	8
Epb4.1	60	0	0	0	1	0	61
Gabrb3	35	3	0	0	0	147	185
Gnai2	29	1	2	3	0	0	35
Gpc	1	0	0	0	0	0	1
Gprasp1	0	0	2	1	1	0	4
Gtse1	13	2	1	3	0	40	59
Il17re	1	0	0	0	0	2	3
Inpp5f	12	0	0	0	0	32	44
Klkl12	47	1	4	2	0	0	54
Mkrn1	0	0	0	0	0	2	2
Mtr	0	0	0	0	0	2	2
Ncoa1	0	0	0	0	0	6	6
Nploc4	0	0	0	0	0	1	1
Nup205	2	0	0	0	0	70	72
Oxr1	94	1	2	0	2	125	224
Pcp4l1	49	4	1	0	0	5	59
Pdia6	6	0	2	0	0	34	42
Pole	6	0	0	3	1	9	19
Prpf40a	45	0	0	0	0	3	48
Pvrl2	8	0	0	0	0	22	30
Pxk	32	0	3	2	1	38	76
Rab3gap2	0	0	0	0	0	35	35
Rad17	0	0	0	0	0	1	1
Rapgef3	0	0	0	0	0	1	1
Rcc2	50	0	0	3	0	0	53
Hps4	2	0	0	0	0	0	2
Rpl5	6	0	0	0	0	0	6
Rps19	0	0	0	0	0	1	1
Samd4b	6	0	0	0	0	3	9
Sema6d	16	1	17	2	4	0	40
Son	0	0	2	0	0	0	2
Spcs2	3	0	0	0	0	2	5
2610019P18Rik	1	0	1	1	0	2	5
Tm9sf2	5	0	0	0	0	6	11
Tmem11	24	2	0	0	0	0	26
Tmem39a	1	0	0	0	0	0	1
Tns3	22	0	0	0	0	0	22
Tspan5	162	0	3	1	0	0	166
Ybx1	72	1	66	1	1	0	141
Yy1	0	0	0	0	0	2	2

Table 4.4: Detailed SNP information. Cn, coding nonsynonomous; Cs, coding synonomous; UTR, untranslated region; I, intron (except first and last two bases); and L, locus region (a region that is not transcribed but may be a regulatory region for the gene locus). The shading highlights instances of each kind of SNP.

Transcription Factor Binding Site Overrepresentation Analysis

Transcription factor binding site (TFBS) analysis was used to test for a common element controlling the expression of the OD genes. The region spanning 2,000 base pairs prior to the transcription start site of the overdominant and underdominant genes was scanned for the presence of TFBSs using the oPOSSUM database (Ho Sui, Mortimer et al. 2005). Five transcription factors were identified that had a significant p-value and/or Z-score for overrepresentation (table 4.5). Thirty-seven genes with overdominant expression were analyzed for TFBS overrepresentation and the TFBS for ELK1 and SP1 were significantly represented in 23 and 20 genes respectively. Eighteen overdominant genes shared TFBS for ELK1 and SP1 (the consensus sites for ELK1 and SP1 contain distinct sequences) indicating a level of higher order transcriptional regulation. Forty-two genes with underdominant expression in the F1 were included in the overrepresentation analysis. The TFBS for *Roaz*, RREB1 and Pax4 were found to be significantly overrepresented amongst the underdominant genes. Relatively few underdominant genes contained the TFBS for RREB1 (4 genes) and Pax4 (2 genes) but the p-values and Z-scores for overrepresentation of these two TFBS were highly significant. Twelve genes contained the TFBS for *Roaz*. No underdominant genes contained all three TFBS, but three genes contained the TFBS for both *Roaz* and RREB1 and two genes contained the TFBS for both RREB1 and *Pax4*.

Overdominance Genes: Over-expressed

TF	TF Class	TF Supergroup	IC	Included/ Excluded	Background gene hits	Background gene non-hits	Target gene hits	Target gene non-hits	Background TFBS hits	Background TFBS rate	Target TFBS hits	Target TFBS rate	Z-score	Fisher score	Genes
ELK1	ETS	vertebrate	8.812	37/6	6812	8338	23	14	15290	0.0222	51	0.0281	5.383	2.67E-02	Tubb5, Pole,**Ascl1**, **Pex13**, **Tns3**, Sgpp1, **Yy1**, **Oxr1**, **Pvrl3**, **Pa2g4**, **Olfm1**, **Tspan5**, **Epb4.1**, **Gnb2**, **Adm**, Rsf1, **Pcp4l1**, Nploc4, Ghitm, **Il17re**, **Esf1**, **Chd8**, **Prpf40a**
SP1	ZN-FINGER, C2H2	vertebrate	9.719	37/6	6154	8996	20	17	15402	0.0223	63	0.0347	11.24	6.87E-02	Tubb5, **Ascl1**, **Pex13**, **Tns3**, Elac2, **Yy1**, **Oxr1**, **Pvrl3**, **Pa2g4**, **Olfm1**, **Tspan5**, **Epb4.1**, **Gnb2**, Emp1, **Adm**, **Pcp4l1**, **Il17re**, **Esf1**, **Chd8**, **Prpf40a**

Overdominance Genes: Under-expressed

TF	TF Class	TF Supergroup	IC	Included/ Excluded	Background gene hits	Background gene non-hits	Target gene hits	Target gene non-hits	Background TFBS hits	Background TFBS rate	Target TFBS hits	Target TFBS rate	Z-score	Fisher score	Genes
Roaz	ZN-FINGER, C2H2	vertebrate	17.925	42/9	1991	13159	12	30	2600	0.0057	13	0.0067	2.343	6.52E-03	Pgrmc1, C330007P06Rik, Ncoa1, Rapgef3, *Add3* , Arfip2, Cryab, Gpc1, Samd4b, *Rcc2* , Ctxn1, *Anks1b*
RREB1	ZN-FINGER, C2H2	vertebrate	22.278	42/9	307	14843	4	38	333	0.001	5	0.0034	13.48	1.05E-02	Add3, Erf, *Rcc2* , *Anks1b*
Pax4	PAIRED-HOMEO	vertebrate	11.004	42/9	62	15088	2	40	63	0.0003	2	0.0021	18.25	1.35E-02	Pdia6, *Erf*

Table 4.5: Results of oPOSSUM Transcription Factor Binding Site Analysis (TFBS). TF is transcription factor, IC is the information content (the specifity of the position weight matrix of the TFBS), Z-score is expressed in magnitude of the standard deviation and takes into account the number of times the TFBS occurs within a target gene list compared to a background gene list and the Fisher score is the p-value for overrepresentation and takes into account the number of genes in a target gene list that contain a given TFBS compared to the number of genes in a background list that contain the same TFBS. Bold indicates overlap between TFBS in genes over-expressed in the F1 (top panel) and italics represent overlap between TFBS in genes under-expressed in the F1 (bottom panel).

Several genes with overdominant expression overlap with genes expressed in high alcohol preferring mice and are alcohol responsive

In chapter 2, the brain gene expression of three selected lines of mice and six isogenic strains of mice that differed in their consumption of alcohol were compared and gene expression that was shared between high and low consuming mice was identified (Mulligan, Ponomarev et al. 2006). The F1 mice in the present study were included in the larger panel of mice from the previous study. In order to asses if any of the OD genes are genes that are known to have divergent expression between high and low alcohol consuming mice, the ODgenes were overlapped with the results of the meta-analysis in chapter 2. Only 4 genes that showed F1 underdominance overlapped with low gene expression in high drinking mice compared to low drinking mice and 9 genes showed overdominant expression in the F1 compared to the parental B6 and FVB genotypes and had high expression in high drinking mice compared to low drinking mice (table 4.6A). *Dab2* (Disabled homolog 2) had also been previously identified as a vulnerability loci for alcohol dependence in humans through a pooled association genome scan using SNPs in populations of alcoholics and non-alcoholics (Johnson, Drgon et al. 2006).

Several microarray studies completed at the University of Texas (S.E. Bergeson) involved B6.FVB hybrid animals that had been exposed to alcohol using the DID paradigm. In the DID procedure mice are given access to a solution of alcohol for a limited period during the first few hours of the dark cycle. Mice are nocturnal animals and they tend to be more active in the initial phase of the dark cycle and will tend to consume more fluid during this period of time. When given alcohol solutions as their only source of liquid mice tend to consume moderate to large quantities of alcohol in a short time frame and this method has proven useful for generating high BACs in mice

(Rhodes, Best et al. 2005). B6 mice consume a great deal of alcohol using this paradigm and F1 mice consume comparable levels (Blednov, Metten et al. 2005; Rhodes, Ford et al. 2007). Whole brain gene expression of adult male FVB.B6 mice that had been allowed to consume alcohol according to the DID paradigm (Rhodes, Best et al. 2005) was compared to control alcohol naïve mice. The F1 mice used in this study had a mean BAC of 0.9 mg/ml ±2 (SEM) (0.09%) which is actually above the legal limit for alcohol intoxication for humans observed in most American states. Eleven overdominant genes showed a significant change in expression after alcohol consumption (table 4.6B). All of the genes that were alcohol responsive were genes whose expression was above that of the parental strains (overdominant). Interestingly, about half of the overdominant genes showed increased expression after alcohol consumption in the DID paradigm and the remaining half showed decreased expression. Four of the ethanol responsive overdominant genes were also identified previously in the meta-analysis overlap. It should be noted here that the animals used in the present study were female animals and that the animals that were exposed to alcohol were male animals. Female mice usually consume more alcohol than male mice, however, both male and female F1 hybrids consumed more alcohol than their same sex parents (Blednov, Metten et al. 2005). The mice in the DID study achieved relatively high BACs and the gene expression changes observed for the overdominant genes are most likely the result of alcohol administration so we can hypothesize that the overdominant genes may behave the same in female animals when exposed to the same level of alcohol. Further testing will be required to validate these assumptions.

A

Overdominance Genes

CLID	GBA	EntrezID	Name	Symbol	B6 AVG	FVB AVG	F1 AVG	B6vsFVB p (HSD)	B6vsFVB FC	F1vsB6p (HSD)	F1vsB6 FC	F1vsFVB p (HSD)	F1vsFVB FC	
H3002C06	BG076418	68626	ElaC homolog 2 (E. coli)	Elac2	5.37	5.36	6.14	9.98E-01	-1.01	2.29E-02	-1.71	2.06E-02	-1.72	
H4070C09	BQ561431	240756	Kelch-like 12 (Drosophila)	Klhl12	4.99	5.02	5.74	9.93E-01	1.02	1.94E-02	-1.67	2.37E-02	-1.64	
AI846740	AI846740	13481	Dolichol-phosphate (beta-D) mannosyltransferase 2	Dpm2	8.57	8.68	9.23	7.68E-01	1.08	4.67E-03	-1.58	1.63E-02	-1.46	
AI839645	AI839645	18813	Proliferation-associated 2G4	Pa2g4	8.00	8.11	8.46	6.61E-01	1.08	9.79E-03	-1.38	4.75E-02	-1.27	
H3094B06	BG071049	20658	Son cell proliferation protein	Son	8.37	8.27	8.82	7.89E-01	-1.07	3.03E-02	-1.36	9.26E-03	-1.46	
H3040G07	BG066254	11750	Annexin A7	Anxa7	6.88	6.84	7.21	9.47E-01	-1.03	4.91E-02	-1.26	2.80E-02	-1.29	
AI835654	AI835654	66425	Purkinje cell protein 4-like 1	Pcp4l1	9.73	9.74	9.99	1.00E+00	1.00	4.55E-02	-1.19	4.57E-02	-1.19	
H3126H03	BG073680	216150	Cell division cycle 34 homolog (S. cerevisiae)	Cdc34	10.44	10.52	10.66	2.93E-01	1.06	2.30E-03	-1.17	3.78E-02	-1.11	
AI837752	AI837752	56177	Olfactomedin 1	Olfm1	11.55	11.57	11.77	9.71E-01	1.01	1.56E-02	-1.16	2.37E-02	-1.15	
H3121G02	BG073362	68035	RIKEN cDNA 3100004P22 gene	3100004P22Rik	9.76	9.64	9.38	3.39E-01	-1.08	1.36E-03	1.29	1.80E-02	1.19	
583640	AI450544	56790	DNA segment, Chr 3, ERATO Doi 300, expressed	D3Ertd300e	6.29	6.39	5.74	8.62E-01	1.07	4.09E-02	1.46	1.61E-02	1.57	
H3112G10	BG072606	13132	Disabled homolog 2 (Drosophila)	Dab2	6.20	6.21	5.61	9.98E-01	1.01	6.92E-03	1.50	6.19E-03	1.51	*
H4051E07	BQ558326	101490	Inositol polyphosphate-5-phosphatase F	Inpp5f	6.06	6.18	5.43	8.30E-01	1.09	3.32E-02	1.54	1.16E-02	1.68	

B

Ethanol Responsive Genes

CLID	GB ID	Name	Symbol	DID F1 AVG c	DID F1 AVG e	p-value EvsC	FC EvsC	F1vsB6 FC	F1vsFVB FC	
614891	AI448892	Guanine nucleotide binding protein, beta 2	Gnb2	3.40	6.53	0.011	-8.78	-5.08	-4.85	
AI854533	AI854533	RIKEN cDNA D330023K18 gene	D330023K18Rik	7.43	7.99	0.002	-1.47	-1.54	-2.21	
AI844823	AI844823	Erythrocyte protein band 4.1	Epb4.1	6.74	7.28	0.046	-1.46	-3.73	-3.44	
H3126H03	BG073680	Cell division cycle 34 homolog (S. cerevisiae)	Cdc34	9.62	10.04	0.003	-1.34	-1.17	-1.11	*
AI846740	AI846740	Dolichol-phosphate (beta-D) mannosyltransferase 2	Dpm2	8.38	8.75	0.034	-1.29	-1.58	-1.46	*
AI837752	AI837752	Olfactomedin 1	Olfm1	12.02	12.24	0.020	-1.16	-1.16	-1.15	*
AI851981	AI851981	Growth hormone inducible transmembrane protein	Ghitm	11.71	11.49	0.043	1.16	-1.13	-1.16	
368696	AI415258	Achaete-scute complex homolog-like 1 (Drosophila)	Ascl1	8.58	8.17	0.041	1.33	-1.31	-1.39	
AI836222	AI836222	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	Sec22b	10.01	9.46	0.009	1.46	-1.51	-1.26	
AI839645	AI839645	Proliferation-associated 2G4	Pa2g4	8.84	8.16	0.010	1.61	-1.38	-1.27	*
H3046A07	BG066705			7.53	5.33	0.032	4.61	-1.64	-2.08	

Table 4.6: Significant Overdominant genes known to be associated with alcohol consumption in high alcohol consuming mice and alcohol responsive genes. CLID, clone identification; GBA, Genebank accession; EntrezID, Entrez gene number; AVG, average, HSD, Tukey's post hoc test *p*-value; FC, fold change; blue, B6; green, F1; and yellow, FVB. (A) Most overlaps occur in the direction of enhanced expression in the F1 and high expression in high alcohol consuming animals from the meta-analysis in chapter 2. (B) Alcohol Responsive Genes. DID F1 AVGc and DID F1 AVG e stand for the mean intensity value for the control group and ethanol group respectively. E or e, alcohol; C or c, control; and p-value, from t-test. Columns 7 and 8 contain p-value and fold change information from the t-test comparison between the ethanol and control group. The 9th and 10th columns contain fold change information for the F1 compared to each parental strain for the naïve female animals. An asterisk indicates overlap with the meta-analysis data. Blue shading indicates genes up-regulated by alcohol in F1 male animals and the tan shading indicates genes down-regulated by ethanol in F1 male animals.

DISCUSSION

Despite having been inbred for many generations, B6 and FVB were originally derived from different progenitor strains. B6 was derived from the murine stock of CC little in 1921 and FVB originated from Swiss Webster mice in 1932 (Beck, Lloyd et al. 2000), consequently, the two parental strains are not closely related (Tsang, Sun et al. 2005) and are highly polymorphic (figure 4.8) . According to the Mouse Phenome Database, there are currently, 10,749,600 and 7,504,840 SNPs that have been documented for B6 and FVB respectively (Bogue, Grubb et al. 2007). The existence of SNPs between B6 and FVB parental strains could result in a difference in expression levels or protein function between parental strains which could be transmitted to the F1. The contributions of many allelic differences in the parental strains may underlie the large number of genes that follow the overdominant mode of inheritance in the F1. Overdominant gene expression could be caused by the interaction of polymorphisms in *cis* or by the combination of dominant alleles interacting in *trans*. Regardless of the mechanisms of action, a large number of genes show overdominant expression in the F1 and over 100 are statistically significantly overdominant. Nearly half of these overdominant genes contain known SNPs between the parental strains. Future quantitiave trait loci (QTL) studies may be able to establish if any of the overdominant genes in this study that contain SNPs between the parental strains are linked to the phenotype of alcohol consumption. It will also be interesting to dissect the contributions of dominant and overdominant alleles to the phenotype of high alcohol consumption observed in the FVB.B6 F1 strain.

Based on the results of this study, it appears that the mode of inheritance for genes that are significantly divergent in at least one of the three strains ($p<0.01$, ANOVA) is largely dominant. Perhaps dominant expression is the norm between two highly

divergent inbred strains that contain a large number of dominant alleles. It does seem that a great many B6 and FVB alleles appear to be dominant in one strain and not the other given the large number of highly divergent transcripts between the parental strains. One such study of F1 inheritance patterns in a cross between A/J and B6 mice did not observe a large effect of dominance in the F1, although many instances of overdominance were observed (Cui, Affourtit et al. 2006). However, in the previous study, a different B6 hybrid system was analyzed and gene expression was compared in a different tissue (liver). In addition, both reciprocal crosses were included as well as an estimation of the contributions of technical error to the estimates of heritability. More research is needed to investigate how the expression patterns in other F1 hybrids are influenced by parental genotype when other inbred strains of mice are used.

The main focus of this study is on genes that show overdominant patterns of expression (overdominance and underdominance) in the F1 hybrid of FVB and B6 mice because the phenotype of two-bottle choice alcohol consumption also shows overdominance. Many of the overdominant genes are assigned to diverse functional groups and are involved in several different cellular processes including signal transduction and transcriptional regulation. A change in the levels of these genes could have a profound effect on many other genes or pathways in the brain. Interestingly there is some evidence for coregulation of some of the overdominant genes based on TFBS overrepresentation and chromosomal location. As discussed previously, a number of the overdominant genes contain SNPs between parental strains that may contribute to the overdominance gene expression phenotype of the F1. Further research is needed to determine the molecular genetic basis for the overdominant patterns of inheritance observed in the F1.

Interestingly, thirteen genes that show expression that is significantly enhanced in the F1 compared to the B6 and FVB parental strains overlap with genes that have previously been associated with high alcohol consumption. It is possible that enhancement of these genes can lead to an increase in two-bottle choice alcohol consumption in the F1, although the genes that overlap here are likely to be small effect size genes because the difference in expression of those genes between the high consuming parental strain and the low consuming parental strain is minimal. Currently, studies that compare gene expression before and after alcohol consumption are underway in the F1. An attempt was made, using preliminary data from one such study to identify overdominant genes that are alcohol responsive. As was the case of the 13 genes that overlap with increased expression in the F1 and high alcohol consuming mouse models, all of the genes that showed a response to alcohol were over-expressed in the F1. About half of the alcohol responsive genes showed an increase after alcohol exposure. Of interest is the fact that *Cdc34*, *Dpm2* and *Olfm1* are overexpressed in the F1, increased after drinking and have similar expression in mouse models of high alcohol consumption. *Gnb2* is also an interesting candidate because its expression levels are greatly enhances after alcohol consumption.

Much work is still needed to dissect the genetic components of inheritance in the FVBxB6 F1 hybrid and determine which genes are involved in the overdominant alcohol phenotype observed in the F1. This study offers a first glimpse into the modes of inheritance observed as well as into the genes that share an overdominant expression phenotype with a behavioral phenotype. Future studies, such as QTL analysis, will be needed to confirm the involvement of the overdominant genes detected in this study and the specific functions of each gene in the development of two-bottle choice alcohol drinking can be investigated.

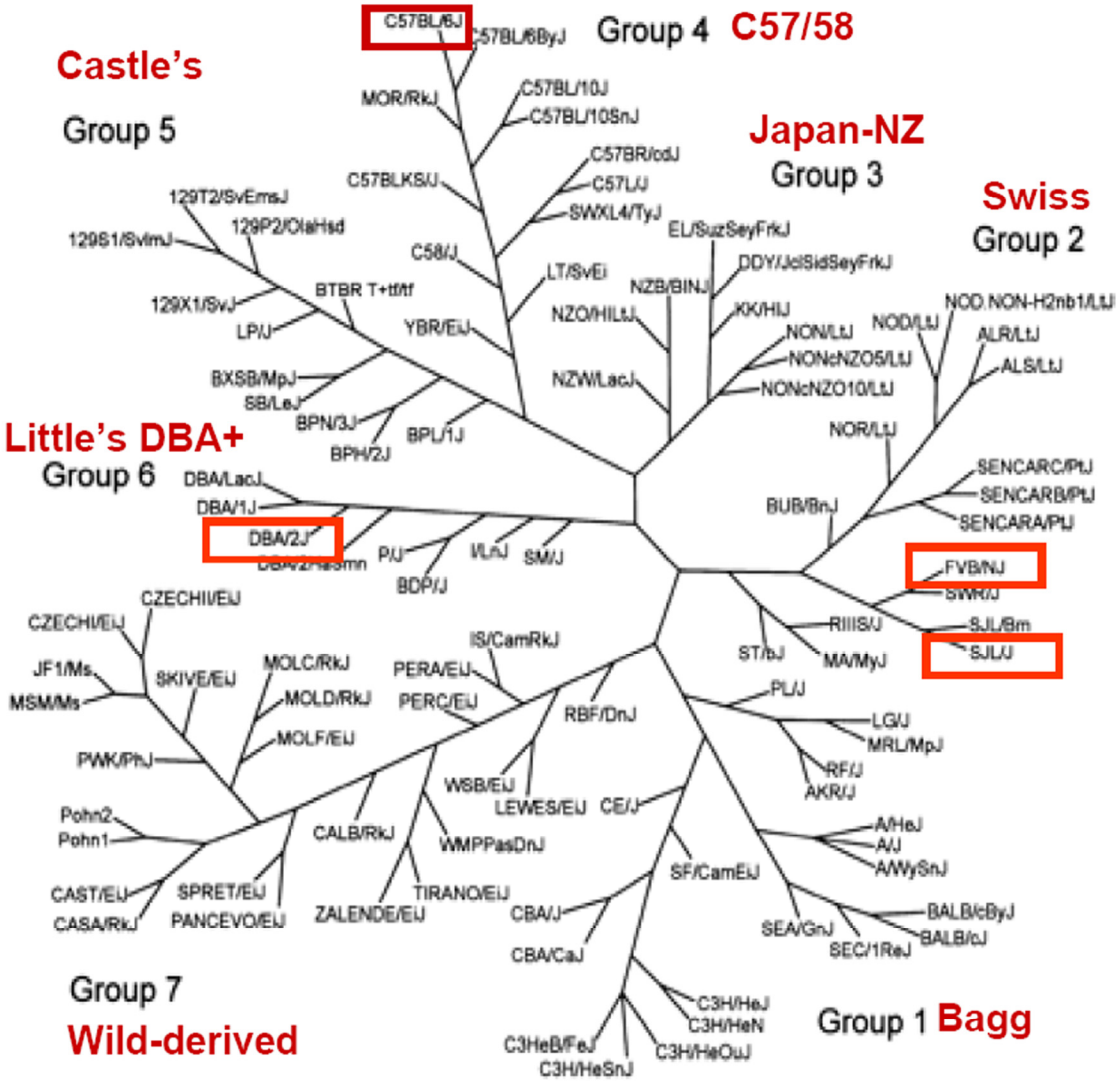


Figure 4.8: Relatedness between inbred mouse strains. Red lettering indicates origin of each cluster. Arm length of the branches gives an indication of relatedness. Red boxes highlight inbred mouse strains discussed in one or more of the chapters (Petkov, Ding et al. 2004).

MATERIALS AND METHODS

Animal husbandry

B6, FVB and FVB.B6 (F1) mice were bred in house from adult male and female mice (1:1) supplied from Jackson Laboratories. FVB, B6 and F1 mice bred in house were weaned at 21 days and fed a diet of PROLAB RMH 1800 which contains a minimum of 18% crude protein, 5% minimum crude fat and 5% minimum crude fiber. Weaned animals were housed with up to 5 same sex animals on a 12 hour light/dark cycle until the mice were sacrificed for microarray analysis at 70 to 100 days old. All animals were naïve to alcohol.

Microarray analysis

Tissue collection, RNA extraction and hybridization was completed as described in the methods section of chapter 3 with the exception that the whole brain (excluding olfactory bulbs) from 5 female animals of each strain was included in the microarray analysis.

Determination of additive and dominance effects

For each trait (transcript), the value of a, the additive effect or the average effect of an allele substitution (Falconer and Mackay 1996), was calculated as one-half the phenotypic (expression level) difference between the means of the two homozygous inbred strains, or B6 and FVB. In this case, the difference was always calculated as B6 – FVB, which yields positive values of *a* when the B6 genotype shows higher expression values, and negative if the FVB genotype shows higher expression values. Also, *d*, the dominance deviation (Falconer and Mackay 1996; Kearsey and Pooni 1996), was calculated as the difference between the phenotypic mean of the F1 heterozygotes and the

average (midpoint) of the two inbred strains. Both *a* (additive) and *d* (dominance) effects on expression values were expressed in pooled within genotype SD units. The sign of *d* was positive if the F1 heterozygote mean expression level scored above the mean of the two inbred strains, and was negative if below. The ratio *d*/*a* was then determined (Kearsey and Pooni 1996); this value is 0 with no dominance, 1.0 with complete B6 allele dominance, and >1.0 with B6 allele overdominance. On the other hand, negative *d/a* ratios reflects FVB dominance or overdominance, respectively. The sign of *d/a* was positive if the B6 allele showed any degree of dominance (FVB allele is recessive), but was negative if the FVB allele showed any degree of dominance (B6 allele is recessive). These parameter estimates were calculated using a custom *R* code script file (courtesy of J.K. Belknap, script available on request) (Storey and Tibshirani 2003).

Tests of significance for a (additive) and d (dominance) effects on expression values.

The observed values of *a* and *d* were tested against the null hypothesis that either parameter was equal to zero. Because *a* and *d* were expressed in pooled within genotype standard deviation units, the following equation was used to calculate Student's *t* for a two-group, two-tailed comparison (Equation 1): $t = |a| \ (df)^{1/2}$, where $df = 8$ for *a* (Rosenthal 1994). For *d*, |*d*| was substituted for |*a*| in Equation 1 and $df = N$ (15 arrays) – 2 = 13 for a test of the F1 mean tested against the mean of both inbred strains pooled together. The test for overdominance was a test that the F1 mean fell outside the range of the two inbred strains and was significantly different from each of them at $p < 0.05$ (Cui, Affourtit et al. 2006). Note that this does not correct for multiple testing. Histograms, correlations and scatterplots based on calculation of additive and dominance effects and significance were graphed using the R 2.6.0 software for Microsoft Windows®.

Hierarchical and K-means clustering

Hierarchical clustering by experiment was performed using the Longhorn Array Database (LAD) on centered log2 R/G median values (Killion, Sherlock et al. 2003). Only transcripts that were significantly different by ANOVA (P<0.01) were clustered. K-means clustering (k=6) was performed on centered and normalized log2 R/G median values using the freely distributed Cluster 3.0 (de Hoon, Imoto et al. 2004) Raster displays were created with JavaTreeview (Saldanha 2004) software.

Functional Overrepresentation Analysis

Functional group/pathway analysis was performed using the DAVID Bioinformatics Resource (http://david.abcc.ncifcrf.gov/) and the WebGestalt Gene Set Analysis Toolkit (Zhang, Kirov et al. 2005). Transcription factor binding site (TFBS) overrepresentation analysis was completed using the oPOSSUM database to analyze 2,000 bp upstream of the transcription factor start site for overdominant and underdominant gene lists as previously described in the methods section of chapter 3 (Ho Sui, Mortimer et al. 2005).

Chromosome distribution and SNP analysis of overdominant genes

Overdominant (overdominant and underdominant) genes were plotted according to their approximates chromosome location using the chromosome distribution chart function in the WebGestalt database (Zhang, Kirov et al. 2005). SNP frequencies and descriptions were extracted from the Mouse Phenome Database (MPD, http://phenome.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home) SNP collection.

Overlap with the alcohol preferring mouse models from the meta-analysis

Transcripts that showed significant overdominant expression (see methods section above) in the F1 were compared to transcripts that were expressed significantly higher (Q<0.05 and average $|d|$>0.5) in alcohol preferring mouse models compared to alcohol non-preferring mouse models. Transcripts that showed underdominant expression in the F1 were compared to transcripts that were expressed significantly lower ((Q<0.05 and average $|d|$>0.5) in alcohol preferring models compared to non-preferring models.

Chapter 5: Conclusion

MICROARRAY META-ANALYSIS

Much evidence from studies in humans and animals supports the hypothesis that alcohol addiction is a complex disease with both hereditary and environmental influences. The contribution of many environmental factors makes it difficult to study the genetic determinants of excessive alcohol consumption in humans. Fortuitously, many mouse models show high, or low, voluntary alcohol drinking providing a unique opportunity to approach the molecular complexities underlying the genetic predisposition to drink alcohol. Three different microarray studies were completed in order to characterize the brain expression profiles of many of these mouse models.

In the first study, microarray analyses of brain gene expression in 3 selected lines and 6 isogenic strains of mice known to differ markedly in voluntary alcohol consumption provided over 4.5 million data points for a meta-analysis. A total of 107 microarrays were obtained and arranged into 4 experimental data sets allowing identification of 3,800 unique genes significantly and consistently changed between all high and low alcohol consuming models. Seventy-five top candidate genes were identified including some genes known to be involved in alcohol consumption or associated with a response to drugs of abuse such as *Prkce* and *Gnb1* respectively. Other top candidate genes, such as *B2m* and *Scn4b*, have not previously been associated with alcohol or addiction phenotypes and may not otherwise have been identified as playing a role in alcohol consumption in mice. Preliminary results from several other laboratories indicate that *B2m* and *Scn4b* are indeed involved in several responses to alcohol (unpublished data). The 3,800 unique transcripts consistently changed in the brains of high and low alcohol consuming mouse models were further filtered into 36 genes that

overlap with mouse alcohol preference QTL and human QTL for alcohol vulnerability. Further analysis may reveal if these genes are involved in aspects of alcohol consumption in mice and the genetic predisposition to alcoholism in human populations. Data from the general meta-analysis was further filtered by a congenic strain microarray set, from which *cis*-regulated candidate genes for an alcohol preference quantitative trait locus on chromosome 9 were identified: *Arhgef12, Carm1, Cryab, Cox5a, Dlat, Fxyd6, Limd1, Nicn1, Nmnat3, Pknox2, Rbp1, Sc5d, Scn4b, Tcf12, Vps11, Zfp291* and four expressed sequence tags. Four of these genes, including *Scnb4* are also correlated with alcohol preference. The research described here provides the first use of a microarray meta-analysis to analyze expression associated with a particular alcohol related behavior, in this case, alcohol preference, and the first use of a congenic strain for the identification of *cis*-regulated transcripts. We also introduce new methodologies for comparing microarray experiments across different platforms and different laboratories. Our microarray meta-analysis is an important database of gene expression differences between mouse models divergent in alcohol consumption and is an extremely valuable tool for all investigators in the field of alcohol and addiction research.

BRAIN REGION MICROARRAY ANALYSIS BETWEEN C57BL/6 SUBSTRAINS

Owing to their predisposition for high voluntary alcohol consumption, C57BL/6 mice have been widely used as a research model; including for the study of alcohol-related traits. In the second microarray study, two substrains of B6 mice separated by a span of approximately 50 years at two different breeding facilities, Charles River Laboratories (C57BL/6NCrl) and Jackson Laboratories (C57BL/6J), showed a significant difference in both two-bottle choice alcohol consumption and preference. The divergence in drinking behavior is likely not due to environmental differences, divergence in taste discrimination or metabolic differences between the substrains. Microarray analysis was used to determine strain differences in brain gene expression for the cortex and striatum, hippocampus, cerebellum and the ventral brain region. Although C57BL/6J and C57BL/6NCrl mice share the same B6 name, evidence was provided that they differ fundamentally at the level of the transcriptome as we found 86 transcripts that are consistently and highly changed between substrains across all four brain regions and over 100 transcripts that are highly significantly changed in one or more brain regions. Genome-wide analysis for clusters of significantly co-localized and co-regulated genes revealed 4 chromosomal regions that contain clusters of genes that are regulated by higher order mechanisms. Analysis of chromosomal distribution also revealed regions on chromosomes 14 and 4 that contain apparent copy number gains in the C57BL/6J genome that may be missing in the C57BL/6NCrl substrain. Interestingly, the expression of a representative gene from the chromosome 14 region, *D14Ertd449e*, seems to vary between inbred strains of mice and data from several comparative genomic scans seems to support the hypothesis that different copy numbers exist for this gene in different inbred strains as well as between the B6 substrains and could account for the observed differences in brain gene expression. One of the overarching goals of the research

presented here is to integrate various microarray databases in order to facilitate the identification of genes involved in alcohol consumption in mice. Towards that goal, gene expression between the C57BL/6 substrains was compared to gene expression between high and low alcohol consuming mouse models from the results of the microarray meta-analysis in chapter 2 and 24 genes were identified that overlapped between datasets. The results of this analysis seem to indicate that, not surprisingly, C57BL/6J mice have retained more genes/transcripts associated high alcohol consumption and preference as evidenced by the fact that the majority of the overlap with preferring animals occurs in C57BL/6J mice. The small amount of genes shared between C57BL/6NCrl mice and alcohol non-preferring mouse models is consistent with the observation that the Charles River Laboratory substrain has an intermediate alcohol consumption phenotype. Although not included as one of the 3,800 genes consistently changed between mouse models in the microarray meta-analysis, *D14Ertd449e* is also consistently divergent between high and low alcohol consuming mouse models (d = 1.7), although the false discovery rate estimation was sub-threshold for inclusion into the larger list of 3,800 unique transcripts (q = 0.08). Intriguingly, *D14Ertd449e* expression is also significantly ($p < 0.05$) positively correlated with alcohol preference according to the WebQTL database (www.genenetwork.org). The function of the *D14Ertd449e* gene is unknown but this gene represents a top candidate gene involved in alcohol consumption differences between B6 substrains as well as between mouse models with divergent levels of alcohol consumption and will be the focus of future research. The transcriptome characterization described here represents a valuable tool for the future investigation of phenotypic differences, both alcohol related and otherwise, observed between these sister substrains.

MICROARRAY ANALYSIS OF GENE EXPRESSION IN A FVB.B6 F1 HYBRID

In the field of agriculture, it has long been observed that the offspring resulting from the cross between two different inbred strains commonly show hybrid vigor or overdominance; enhanced agronomic performance compared to either inbred parental strain. In the third microarray study, the phenotype of overdominant two-bottle choice voluntary alcohol consumption was investigated through comparison of the expression patterns in the brains of two inbred parental strains (B6 and FVB) and the resulting F1 hybrid (FVB.B6). The pattern of drinking observed in the F1 was very unusual because most mice tend to show avoidance for highly concentrated alcohol solutions but the F1 will consume large amounts of alcohol at high concentrations. It was previously hypothesized that the enhanced consumption of the F1 hybrid strain could be due to overdominance (Blednov, Metten et al. 2005). Indeed, the results of the expression analysis confirmed that many genes showed overdominant patterns of expression in the F1. It also appeared that most of the significant expression in the F1 showed extensive patterns of dominance. Both overdominant and underdominant genes were involved in many different pathways including transcriptional regulation, cellular signaling and metabolism. Several overdominant genes were identified as being similarly expressed in high or low alcohol consuming animal models, *Cdc34, Dpm2 and Olfm1,* and several genes were very ethanol responsive, including *Gnb2*. These genes are good candidate genes for underlying the enhanced drinking observed in the F1. Interestingly, there was more overlap between genes with overdominant expression (enhanced expression in the F1 compared to either parental strain) and genes expressed at a higher level in alcohol preferring animals from the meta-analysis in chapter 2. No genes with underdominant patterns of expression were found to be alcohol responsive. It is possible that several overdominant genes may be playing a role in reducing avoidance for high concentrations

of alcohol in the F1 and further study is necessary to correlate the expression pattern of the overdominant genes with the overdominant phenotype of high voluntary alcohol consumption. FVBxB6 recombinant inbred strains are currently being generated at another university for the detection of QTL regions for alcohol consumption in the F1 model system and any of the overdominant genes detected by this microarray analysis that overlap with a QTL region will be strong candidate genes for high alcohol consumption. The FVB.B6 F1 hybrid also represents an excellent model system for dissecting the molecular mechanisms that underlie overdominant patterns of expression and investigation of patterns of overdominance in the FVBxB6 recombinant inbred strains should also facilitate identification of the genetic interactions that determine overdominance.

IMPACT

The work described in this dissertation provides a foundation for future research that will eventually contribute to a better understanding of the genetic determinants of alcohol consumption in mice and the genetic vulnerability for addiction in humans. The meta-analytic approach described in chapter 2 is evidence of the power that can be achieved by combining many carefully designed microarray data sets. Our understanding about the genetic underpinnings of alcoholism is only at the tip of the iceberg. Each of the data sets described in detail here has generated hundreds of candidate genes and many hypothesis that can be tested. Through integration of microarray datasets and other bioinformatics resources now and in the future we can isolate candidate genes for high alcohol consumption in mice both within each microarray dataset and across microarray datasets and design experiments that will increase our understanding of the genetic predisposition towards alcoholism in humans.

References

Agabio, R., G. Cortis, et al. (1996). Circadian drinking pattern of Sardinian alcohol-preferring rats. Alcohol Alcohol 31(4): 385-8.

Agrawal, A., A. L. Hinrichs, et al. (2007). Linkage scan for quantitative traits identifies new regions of interest for substance dependence in the Collaborative Study on the Genetics of Alcoholism (COGA) sample. Drug Alcohol Depend.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders : DSM-IV. Washington, DC, American Psychiatric Association.

Barr, C. S. and D. Goldman (2006). Non-human primate models of inheritance vulnerability to alcohol use disorders. Addict Biol 11(3-4): 374-85.

Barrett, T., D. B. Troup, et al. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res 35(Database issue): D760-5.

Beck, J. A., S. Lloyd, et al. (2000). Genealogies of mouse inbred strains. Nat Genet 24(1): 23-5.

Begleiter H, R. T., Hesselbrock V, Porjesz B, Li T-K, Schuckit MA, Edenberg HJ, Rice AP (1995). The collaborative study on the genetics of alcoholism. Alcohol Health Res World 19: 228-236.

Behm A, L. L., Li T-K, Grahame N Selective breeding for replicate line 2 high and low alcohol preferring mice. Alcohol Clin Exp Res 27:49A. (2003).

Belknap, J. K. and A. L. Atkins (2001). The replicability of QTLs for murine alcohol preference drinking behavior across eight independent studies. Mamm Genome 12(12): 893-9.

Belknap, J. K., J. C. Crabbe, et al. (1993). Voluntary consumption of ethanol in 15 inbred mouse strains. Psychopharmacology (Berl) 112(4): 503-10.

Belknap, J. K., S. P. Richards, et al. (1997). Short-term selective breeding as a tool for QTL mapping: ethanol preference drinking in mice. Behav Genet 27(1): 55-66.

Bell, R. L., Z. A. Rodd, et al. (2006). The alcohol-preferring P rat and animal models of excessive alcohol drinking. Addict Biol 11(3-4): 270-88.

Bice, P., T. Foroud, et al. (1998). Genomic screen for QTLs underlying alcohol consumption in the P and NP rat lines. Mamm Genome 9(12): 949-55.

Blednov, Y. A., S. E. Bergeson, et al. (2005). Perturbation of chemokine networks by gene deletion alters the reinforcing actions of ethanol. Behav Brain Res 165(1): 110-25.

Blednov, Y. A., P. Metten, et al. (2005). Hybrid C57BL/6J x FVB/NJ mice drink more alcohol than do C57BL/6J mice. Alcohol Clin Exp Res 29(11): 1949-58.

Blum, K., A. H. Briggs, et al. (1982). Whole brain methionine-enkephalin of ethanol-avoiding and ethanol-preferring c57BL mice. Experientia 38(12): 1469-70.

Bogue, M. A., S. C. Grubb, et al. (2007). Mouse Phenome Database (MPD). Nucleic Acids Res 35(Database issue): D643-9.

Campbell, A. D., R. R. Kohl, et al. (1996). Serotonin-3 receptor and ethanol-stimulated somatodendritic dopamine release. Alcohol 13(6): 569-74.

Cano, M., L. L. Arnold, et al. (2001). Evaluation of diet and dimethylarsinic acid on the urothelium of Syrian golden hamsters. Toxicol Pathol 29(6): 600-6.

Cardoso, R. A., S. J. Brozowski, et al. (1999). Effects of ethanol on recombinant human neuronal nicotinic acetylcholine receptors expressed in Xenopus oocytes. J Pharmacol Exp Ther 289(2): 774-80.

Carr, L. G., K. Habegger, et al. (2003). Analyses of quantitative trait loci contributing to alcohol preference in HAD1/LAD1 and HAD2/LAD2 rats. Alcohol Clin Exp Res 27(11): 1710-7.

Coffer, P. J. and B. M. Burgering (2004). Forkhead-box transcription factors and their role in the immune system. Nat Rev Immunol 4(11): 889-99.

Colombo, G., R. Agabio, et al. (1997). Sardinian alcohol-preferring rats prefer chocolate and sucrose over ethanol. Alcohol 14(6): 611-5.

Corl, A. B., A. R. Rodan, et al. (2005). Insulin signaling in the nervous system regulates ethanol intoxication in Drosophila melanogaster. Nat Neurosci 8(1): 18-9.

Council, N. R. (1996). Guide for the Care and Use of Laboratory Animals. Washington, DC, Nat. Res. Council.

Cowen, M. S., F. Chen, et al. (2004). Neuropeptides: implications for alcoholism. J Neurochem 89(2): 273-85.

Crabbe, J. C. (2002). Alcohol and genetics: new models. Am J Med Genet 114(8): 969-74.

Crabbe, J. C., T. J. Phillips, et al. (1999). Identifying genes for alcohol and drug sensitivity: recent progress and future directions. Trends Neurosci 22(4): 173-9.

Crabbe, J. C., T. J. Phillips, et al. (2006). Alcohol-related genes: contributions from studies with genetically engineered mice. Addict Biol 11(3-4): 195-269.

Crabbe, J. C., T. J. Phillips, et al. (1990). Estimation of genetic correlation: interpretation of experiments using selectively bred and inbred animals. Alcohol Clin Exp Res 14(2): 141-51.

Cui, X., J. Affourtit, et al. (2006). Inheritance patterns of transcript levels in F1 hybrid mice. Genetics 174(2): 627-37.

Dackis, C. and C. O'Brien (2005). Neurobiology of addiction: treatment and public policy ramifications. Nat Neurosci 8(11): 1431-6.

Dawson, D. A. and B. F. Grant (1998). Family history of alcoholism and gender: their combined effects on DSM-IV alcohol dependence and major depression. J Stud Alcohol 59(1): 97-106.

de Hoon, M. J., S. Imoto, et al. (2004). Open source clustering software. Bioinformatics 20(9): 1453-4.

DeRisi, J. L., V. R. Iyer, et al. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278(5338): 680-6.

Devoto, P., G. Colombo, et al. (1998). Serotonin is reduced in the frontal cortex of Sardinian ethanol-preferring rats. Alcohol Alcohol 33(3): 226-9.

Diehn, M., G. Sherlock, et al. (2003). SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res 31(1): 219-23.

Duggan, D. J., M. Bittner, et al. (1999). Expression profiling using cDNA microarrays. Nat Genet 21(1 Suppl): 10-4.

Eckardt, M. J., S. E. File, et al. (1998). Effects of moderate alcohol consumption on the central nervous system. Alcohol Clin Exp Res 22(5): 998-1040.

Engracia, V., M. M. Leite, et al. (2003). Expression of class mu glutathione-S-transferase in human liver and its association with hepatopathies. Am J Med Genet A 123(3): 257-60.

Enoch, M. A. (2003). Pharmacogenomics of alcohol response and addiction. Am J Pharmacogenomics 3(4): 217-32.

Eriksson, K. (1968). Genetic Selection for Voluntary Alcohol Consumption in the Albino Rat. Science 159(3816): 739-741.

Falconer, D. S. and T. F. C. Mackay (1996). Introduction to quantitative genetics. Essex, England, Longman.

Finn, D. A., J. K. Belknap, et al. (2005). A procedure to produce high alcohol intake in mice. Psychopharmacology (Berl) 178(4): 471-80.

Flatscher-Bader, T. and P. A. Wilce (2006). Chronic smoking and alcoholism change expression of selective genes in the human prefrontal cortex. Alcohol Clin Exp Res 30(5): 908-15.

Foroud, T., H. J. Edenberg, et al. (2000). Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. Alcohol Clin Exp Res 24(7): 933-45.

Fukunaga, K., T. Ishigami, et al. (2005). Transcriptional regulation of neuronal genes and its effect on neural functions: expression and function of forkhead transcription factors in neurons. J Pharmacol Sci 98(3): 205-11.

Fuller, J. L. (1964). Measurement of Alcohol Preference in Genetic Experiments. J Comp Physiol Psychol 57: 85-8.

Gessa, G. L., F. Muntoni, et al. (1985). Low doses of ethanol activate dopaminergic neurons in the ventral tegmental area. Brain Res 348(1): 201-3.

Ghezzi, A., Y. M. Al-Hasan, et al. (2004). slo K(+) channel gene regulation mediates rapid drug tolerance. Proc Natl Acad Sci U S A 101(49): 17276-81.

Gibson, G., R. Riley-Berger, et al. (2004). Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. Genetics 167(4): 1791-9.

Gibson, G. and B. Weir (2005). The quantitative genetics of transcription. Trends Genet 21(11): 616-23.

Goldman, D., G. Oroszi, et al. (2005). The genetics of addictions: uncovering the genes. Nat Rev Genet 6(7): 521-32.

Grahame, N. J., T. K. Li, et al. (1999). Selective breeding for high and low alcohol preference in mice. Behav Genet 29(1): 47-57.

Graubert, T. A., P. Cahan, et al. (2007). A high-resolution map of segmental DNA copy number variation in the mouse genome. PLoS Genet 3(1): e3.

Green, M. L., A. V. Singh, et al. (2007). Reprogramming of genetic networks during initiation of the Fetal Alcohol Syndrome. Dev Dyn 236(2): 613-31.

Grieco, T. M., J. D. Malhotra, et al. (2005). Open-channel block by the cytoplasmic tail of sodium channel beta4 as a mechanism for resurgent sodium current. Neuron 45(2): 233-44.

Hansen, D. (1995). Preventing alcohol abuse: alcohol, culture and control. Westport, Conn., Praeger.

Harwood, H. J., D. Fountain, et al. (1999). Economic cost of alcohol and drug abuse in the United States, 1992: a report. Addiction 94(5): 631-5.

Hiroi, N. and S. Agatsuma (2005). Genetic susceptibility to substance dependence. Mol Psychiatry 10(4): 336-44.

Ho Sui, S. J., J. R. Mortimer, et al. (2005). oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. Nucleic Acids Res 33(10): 3154-64.

Hodge, C. W., K. K. Mehmert, et al. (1999). Supersensitivity to allosteric GABA(A) receptor modulators and alcohol in mice lacking PKCepsilon. Nat Neurosci 2(11): 997-1002.

Hovland, D. N., Jr., R. M. Cantor, et al. (2000). Identification of a murine locus conveying susceptibility to cadmium-induced forelimb malformations. Genomics 63(2): 193-201.

Huang, T. T., M. Naeemuddin, et al. (2006). Genetic modifiers of the phenotype of mice deficient in mitochondrial superoxide dismutase. Hum Mol Genet 15(7): 1187-94.

Huh, G. S., L. M. Boulanger, et al. (2000). Functional requirement for class I MHC in CNS development and plasticity. Science 290(5499): 2155-9.

Iwamoto, K., M. Bundo, et al. (2004). Decreased expression of NEFH and PCP4/PEP19 in the prefrontal cortex of alcoholics. Neurosci Res 49(4): 379-85.

Jeong, H. J., S. H. Hong, et al. (2005). Ethanol induces the production of cytokines via the Ca2+, MAP kinase, HIF-1alpha, and NF-kappaB pathway. Life Sci 77(17): 2179-92.

Johnson, C., T. Drgon, et al. (2006). Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism

vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Genet 141(8): 844-53.

Kearsey, M. J. and H. S. Pooni (1996). The genetical analysis of quantitative traits. London ; New York, Chapman & Hall.

Khisti, R. T., J. Wolstenholme, et al. (2006). Characterization of the ethanol-deprivation effect in substrains of C57BL/6 mice. Alcohol 40(2): 119-26.

Killion, P. J., G. Sherlock, et al. (2003). The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford Microarray Database (SMD). BMC Bioinformatics 4: 32.

Kohnke, M. D. (2007). Approach to the genetics of alcoholism: A review based on pathophysiology. Biochem Pharmacol.

Koob, G. F. (2003). Alcoholism: allostasis and beyond. Alcohol Clin Exp Res 27(2): 232-43.

Kreek, M. J., D. A. Nielsen, et al. (2005). Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. Nat Neurosci 8(11): 1450-7.

Laganiere, J., G. Deblois, et al. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A 102(33): 11651-6.

Le Moal, M. and G. F. Koob (2007). Drug addiction: pathways to the disease and pathophysiological perspectives. Eur Neuropsychopharmacol 17(6-7): 377-93.

Lewohl, J. M., L. Wang, et al. (2000). Gene expression in human alcoholism: microarray analysis of frontal cortex. Alcohol Clin Exp Res 24(12): 1873-82.

Li, T. K., L. Lumeng, et al. (1993). Selective breeding for alcohol preference and associated responses. Behav Genet 23(2): 163-70.

Liang, T., K. Habegger, et al. (2004). Glutathione S-transferase 8-8 expression is lower in alcohol-preferring than in alcohol-nonpreferring rats. Alcohol Clin Exp Res 28(11): 1622-8.

Lippman, Z. B. and D. Zamir (2007). Heterosis: revisiting the magic. Trends Genet 23(2): 60-6.

Lipshutz, R. J., S. P. Fodor, et al. (1999). High density synthetic oligonucleotide arrays. Nat Genet 21(1 Suppl): 20-4.

Liu, J., J. M. Lewohl, et al. (2006). Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. Neuropsychopharmacology 31(7): 1574-82.

Liu, Q. R., T. Drgon, et al. (2006). Addiction molecular genetics: 639,401 SNP whole genome association identifies many cell adhesion genes. Am J Med Genet B Neuropsychiatr Genet 141(8): 918-25.

Lobina, C., R. Agabio, et al. (1997). Constant absolute ethanol intake by Sardinian alcohol-preferring rats independent of ethanol concentrations. Alcohol Alcohol 32(1): 19-22.

Mayfield, R. D., J. M. Lewohl, et al. (2002). Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. J Neurochem 81(4): 802-13.

McClearn, G., Wilson, J. & Meridith, W. (1970). The use of Isogenic and Heterogenic Mouse Stocks in Behavioral Research. New York, Appleton-Century-Crofts.

McClearn, G. E. and D. A. Rodgers (1959). Differences in alcohol preference among inbred strains of mice. Q J Stud Alcohol 20: 691-695.

McClearn GE, W. J., Meredith W (1970). The use of isogenic and heterogenic mouse stocks in behavioral research. New York, Appleton-Century-Crofts.

McClung, C. A. and E. J. Nestler (2003). Regulation of gene expression and cocaine reward by CREB and DeltaFosB. Nat Neurosci 6(11): 1208-15.

McLellan, A. T., D. Carise, et al. (2003). Can the national addiction treatment infrastructure support the public's demand for quality care? J Subst Abuse Treat 25(2): 117-21.

Mereu, G., F. Fadda, et al. (1984). Ethanol stimulates the firing rate of nigral dopaminergic neurons in unanesthetized rats. Brain Res 292(1): 63-9.

Mokdad, A. H., J. S. Marks, et al. (2004). Actual causes of death in the United States, 2000. Jama 291(10): 1238-45.

Moller, C., L. Wiklund, et al. (1997). Decreased measures of experimental anxiety in rats bred for high alcohol preference. Alcohol Clin Exp Res 21(4): 656-60.

Morris, B. J. (2005). A forkhead in the road to longevity: the molecular basis of lifespan becomes clearer. J Hypertens 23(7): 1285-309.

Mulligan, M. K., I. Ponomarev, et al. (2006). Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. Proc Natl Acad Sci U S A 103(16): 6368-73.

Mural, R. J., M. D. Adams, et al. (2002). A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. Science 296(5573): 1661-71.

Nestler, E. J. (2001). Molecular basis of long-term plasticity underlying addiction. Nat Rev Neurosci 2(2): 119-28.

Nestler, E. J. (2005). Is there a common molecular pathway for addiction? Nat Neurosci 8(11): 1445-9.

O'Dell, L. E., A. J. Roberts, et al. (2004). Enhanced alcohol self-administration after intermittent versus continuous alcohol vapor exposure. Alcohol Clin Exp Res 28(11): 1676-82.

Olds, J. and P. Milner (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol 47(6): 419-27.

Olegard, R., K. G. Sabel, et al. (1979). Effects on the child of alcohol abuse during pregnancy. Retrospective and prospective studies. Acta Paediatr Scand Suppl 275: 112-21.

Oswald, L. M. and G. S. Wand (2004). Opioids and alcoholism. Physiol Behav 81(2): 339-58.

Petkov, P. M., Y. Ding, et al. (2004). An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res 14(9): 1806-11.

Phillips, T. J., J. K. Belknap, et al. (1998). Genes on mouse chromosomes 2 and 9 determine variation in ethanol consumption. Mamm Genome 9(12): 936-41.

Phillips, T. J., J. K. Belknap, et al. (2002). Harnessing the mouse to unravel the genetics of human disease. Genes Brain Behav 1(1): 14-26.

Phillips, T. J., J. Broadbent, et al. (2005). Genetic correlational analyses of ethanol reward and aversion phenotypes in short-term selected mouse lines bred for ethanol drinking or ethanol-induced conditioned taste aversion. Behav Neurosci 119(4): 892-910.

Phillips, T. J., J. C. Crabbe, et al. (1994). Localization of genes affecting alcohol drinking in mice. Alcohol Clin Exp Res 18(4): 931-41.

Phillips, T. J., E. S. Terdal, et al. (1990). Response to selection for sensitivity to ethanol hypothermia: genetic analyses. Behav Genet 20(4): 473-80.

Ponomarev, I. and J. C. Crabbe (2002). A novel method to assess initial sensitivity and acute functional tolerance to hypnotic effects of ethanol. J Pharmacol Exp Ther 302(1): 257-63.

Ponomarev, I. and J. C. Crabbe (2004). Characterization of acute functional tolerance to the hypnotic effects of ethanol in mice. Alcohol Clin Exp Res 28(7): 991-7.

Ponomarev, I., R. Maiya, et al. (2006). Transcriptional signatures of cellular plasticity in mice lacking the alpha1 subunit of GABAA receptors. J Neurosci 26(21): 5673-83.

Radulovic, J., J. Kammermeier, et al. (1998). Generalization of fear responses in C57BL/6N mice subjected to one-trial foreground contextual fear conditioning. Behav Brain Res 95(2): 179-89.

Ramachandra, V., S. Phuc, et al. (2007). Ethanol Preference Is Inversely Correlated With Ethanol-Induced Dopamine Release in 2 Substrains of C57BL/6 Mice. Alcohol Clin Exp Res.

Rhodes, D. R., T. R. Barrette, et al. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res 62(15): 4427-33.

Rhodes, J. S., K. Best, et al. (2005). Evaluation of a simple model of ethanol drinking to intoxication in C57BL/6J mice. Physiol Behav 84(1): 53-63.

Rhodes, J. S., M. M. Ford, et al. (2007). Mouse inbred strain differences in ethanol drinking to intoxication. Genes Brain Behav 6(1): 1-18.

Roberts, A. J., C. J. Heyser, et al. (2000). Excessive ethanol drinking following a history of dependence: animal model of allostasis. Neuropsychopharmacology 22(6): 581-94.

Rosenthal, D. (1994). Parametric Measures of Effect Size. New York, Russel Sage Found.

Roth, D. M., J. S. Swaney, et al. (2002). Impact of anesthesia on cardiac function during echocardiography in mice. Am J Physiol Heart Circ Physiol 282(6): H2134-40.

Rothenfluh, A., R. J. Threlkeld, et al. (2006). Distinct behavioral responses to ethanol are regulated by alternate RhoGAP18B isoforms. Cell 127(1): 199-211.

Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics 20(17): 3246-8.

Schena, M., D. Shalon, et al. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270(5235): 467-70.

Scholz, H., M. Franz, et al. (2005). The hangover gene defines a stress pathway required for ethanol tolerance development. Nature 436(7052): 845-7.

Schuckit, M. A. (2000). Genetics of the risk for alcoholism. Am J Addict 9(2): 103-12.

Siegmund, A., K. Langnaese, et al. (2005). Differences in extinction of conditioned fear in C57BL/6 substrains are unrelated to expression of alpha-synuclein. Behav Brain Res 157(2): 291-8.

Sinclair, J. D., A. D. Le, et al. (1989). The AA and ANA rat lines, selected for differences in voluntary alcohol consumption. Experientia 45(9): 798-805.

Snijders, A. M., N. J. Nowak, et al. (2005). Mapping segmental and sequence variations among laboratory mice using BAC array CGH. Genome Res 15(2): 302-11.

Sokolov, B. P., L. Jiang, et al. (2003). Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. J Neurosci Res 72(6): 756-67.

Spanagel, R. (2003). Alcohol addiction research: from animal models to clinics. Best Pract Res Clin Gastroenterol 17(4): 507-18.

Storey, J. D. and R. Tibshirani (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100(16): 9440-5.

Stupar, R. M. and N. M. Springer (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics 173(4): 2199-210.

Swanson-Wagner, R. A., Y. Jia, et al. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci U S A 103(18): 6805-10.

Tabakoff, B., S. V. Bhave, et al. (2005). Why do some mice drink? DNA microarray analysis of candidate genes for alcohol preference. J. Neuroscience In Press.

Tabakoff, B. and P. L. Hoffman (2000). Animal models in alcohol research. Alcohol Res Health 24(2): 77-84.

Tanaka, T. S., S. A. Jaradat, et al. (2000). Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. Proc Natl Acad Sci U S A 97(16): 9127-32.

Tomkins, D. M. and E. M. Sellers (2001). Addiction and the brain: the role of neurotransmitters in the cause and treatment of drug dependence. Cmaj 164(6): 817-21.

Tsang, S., Z. Sun, et al. (2005). A comprehensive SNP-based genetic analysis of inbred mouse strains. Mamm Genome 16(7): 476-80.

Vallee, B. L. (1998). Alcohol in the Western World. Scientific American 278(6): 6.

Wahlsten, D., A. Bachmanov, et al. (2006). Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. Proc Natl Acad Sci U S A 103(44): 16364-9.

Walters, G. D. (2002). The heritability of alcohol abuse and dependence: a meta-analysis of behavior genetic research. Am J Drug Alcohol Abuse 28(3): 557-84.

Wang, J., R. W. Williams, et al. (2003). WebQTL: web-based complex trait analysis. Neuroinformatics 1(4): 299-308.

Wise, R. A. (1998). Drug-activation of brain reward pathways. Drug Alcohol Depend 51(1-2): 13-22.

Wise, R. A. and P. P. Rompre (1989). Brain dopamine and reward. Annu Rev Psychol 40: 191-225.

Workman, C., L. J. Jensen, et al. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol 3(9): research0048.

Worst, T. J. and K. E. Vrana (2005). Alcohol and gene expression in the central nervous system. Alcohol Alcohol 40(1): 63-75.

Yang, Y. H., S. Dudoit, et al. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30(4): e15.

Yao, W. D., R. R. Gainetdinov, et al. (2004). Identification of PSD-95 as a regulator of dopamine-mediated synaptic and behavioral plasticity. Neuron 41(4): 625-38.

Yu, F. H., R. E. Westenbroek, et al. (2003). Sodium channel beta4, a new disulfide-linked auxiliary subunit with similarity to beta2. J Neurosci 23(20): 7577-85.

Zhang, B., S. Kirov, et al. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33(Web Server issue): W741-8.

Zhang, B., D. Schmoyer, et al. (2004). GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. BMC Bioinformatics 5: 16.

Zhang, L., M. F. Miles, et al. (2003). A model of molecular interactions on short oligonucleotide microarrays. Nat Biotechnol 21(7): 818-21.

Vita

Megan Kathleen Mulligan was born in Selma, California on November 28, 1977. She is the daughter of Steve Mulligan and Susan Mulligan and has two younger brothers. Born to parents who had both been raised in farming families, the Mulligans and the Cliftons, she spent much of her early life on a working ranch where she developed an interest in biology. She graduated from Sanger High School in Sanger, California in 1995 and entered the University of California at Santa Cruz (UCSC) in the same year. In 1999 she graduated from UCSC with both a Bachelor of Arts in Biology with highest honors and a Bachelor of Arts in Psychology with honors. Following her graduation, she worked as a Laboratory Technician in the laboratory of Dr. William T. Sullivan at UCSC until she entered the doctoral program in Molecular Biology at the Institute of Cell and Molecular Biology at the University of Texas at Austin in 2001. While a graduate student, she was awarded an F.M. Jones and H.L. Bruce Endowed Graduate Fellowship in Addiction Biology from the Waggoner Center for Alcohol and Addiction Research and received an Honorable Mention Outstanding Teaching Award from the School of Biological Sciences at The University of Texas at Austin. Her research was published in the *Proceedings of the National Academy of Sciences* in April 2006.

Permanent address: 706 South Quality Lane, Sanger, California, 93657

This dissertation was typed by Megan Kathleen Mulligan.